Technical Report 1341

Understanding the Impact of Training on Performance

Beth M. Plott
Patricia L. McDermott
Susan Archer
Thomas F. Carolan
Shaun Hutchins
Alia Fisher
Mark Gronowski
Christopher D. Wickens
Alion Science and Technology

Karin A. Orvis
U.S. Army Research Institute

May 2014

**United States Army Research Institute
for the Behavioral and Social Sciences**

U.S. Army Research Institute
for the Behavioral and Social Sciences

Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:

MICHELLE SAMS, Ph.D.
Director

Research accomplished under contract
for the Department of the Army by:

Alion Science and Technology, Incorporated

Technical review by:

James Belanich, Institute for Defense Analyses
Garett Howardson, George Washington University

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* May 2014	**2. REPORT TYPE** Final	**3. DATES COVERED** *(From - To)* September 2009 to December 2013

4. TITLE AND SUBTITLE Understanding the Impact of Training on Performance	**5a. CONTRACT NUMBER** W91WAW-07-C-0029
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102

6. AUTHOR(S) Beth M. Plott, Patricia L. McDermott, Susan Archer, Thomas F. Carolan, Shaun Hutchins, Alia Fisher, Mark Gronowski, Christopher D. Wickens; Karin A. Orvis	**5d. PROJECT NUMBER** B74F
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER 2902

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Alion Science and Technology, Inc. 4949 Pearl East Circle, #200 Boulder, CO 80301	**8. PERFORMING ORGANIZATION REPORT NUMBER**

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Institute for the Behavioral & Social Sciences 6000 6TH Street (Bldg. 1464 / Mail Stop 5610) Fort Belvoir, VA 22060-5610	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) Technical Report 1341

12. DISTRIBUTION/AVAILABILITY STATEMENT: Distribution Statement A: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Contracting Officer's Representative and Subject Matter POC: Dr. Gerald F. Goodwin

14. ABSTRACT

This research effort was conducted to collect empirical evidence on the effectiveness of different training methods for acquiring and transferring complex cognitive skills. To accomplish this goal, we conducted a series of meta-analyses (and supplemental experiments) examining six training methods (training wheels, scaffolding, part-task training, increasing difficulty, learner control, and exploratory learning), as well as factors that moderate their effectiveness, such as task/skill type being trained (e.g., perceptual, psychomotor, cognitive-declarative), trainee characteristics (e.g., experience, aptitude), and type of training performance outcome (e.g., learning, transfer). Algorithms were developed to quantify the relationships between the training methods, performance, and the various moderating factors. These algorithms can be used to perform tradeoff analyses to determine the effectiveness of different combinations of training method(s), task/skill types, trainee characteristics, and performance outcomes. Finally, to ensure these research findings and algorithms would be easily consumable by training developers and researchers, a training effectiveness tool was developed, called TARGET (which stands for Training Aide: Research and Guidance for Effective Training). This tool can aid training developers and researchers in making evidence-based decisions concerning the most appropriate training method(s) to use depending on their particular training context; thus, helping to maximize effective learning and transfer.

15. SUBJECT TERMS
Transfer of training, Training effectiveness, Meta-analyses, Training methods

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Gerald F. Goodwin
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	Unlimited Unclassified	56	**19b. TELEPHONE NUMBER** (703) 545-2410

Technical Report 1341

Understanding the Impact of Training on Performance

Beth M. Plott
Patricia L. McDermott
Susan Archer
Dr. Thomas F. Carolan
Shaun Hutchins
Alia Fisher
Mark Gronowski
Dr. Christopher D. Wickens
Alion Science and Technology

Karin A. Orvis
U.S. Army Research Institute

Foundational Science Research Unit
Gerald F. Goodwin, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
6000 6th Street, Building 1464
Fort Belvoir, VA 22060

May 2014

UNDERSTANDING THE IMPACT OF TRAINING ON PERFORMANCE

EXECUTIVE SUMMARY

Research Requirement

This research sought to understand how training methods could be effectively employed to improve performance on Army-relevant tasks. We focused on training cognitive skills (e.g., problem solving, quantitative, spatial reasoning, decision making), as these skills are particularly important for developing an adaptable fighting force. The overarching goal of this 4-year program of research was to develop evidence-based guidelines for the effectiveness of six different training methods for acquiring and transferring cognitive skills in complex task domains. We also focused heavily on identifying moderators of a given training method's effectiveness. That is, we were not only interested in identifying whether a method had an overall benefit (or lack thereof), but in identifying for whom specifically that benefit existed and if this benefit varied according to the training performance outcome(s) of interest and type of cognitive task/skill trained.

Procedure

To accomplish this overall goal, we used a combination of literature review, meta-analyses, and experimentation. We conducted a broad literature search to gather evidence on the effectiveness of various training methods. From the literature review, we narrowed the possible training method options to those identified as most suitable for cognitive skills. Based on this literature review, we conducted six comprehensive meta-analyses in order to generate estimates of the effectiveness of the following training methods: (1) training wheels, (2) scaffolding, (3) part-task training, (4) increasing difficulty, (5) exploratory learning, and (6) learner control. The first four methods focused on facilitating cognitive load reduction during learning, while the final two methods focused on increasing active learning and learner engagement during learning. Increasing active learning and learning engagement have both been demonstrated to positively impact training effectiveness.

Second, we conducted five research experiments to help fill several identified research gaps from the meta-analyses. These experiments included several common design elements, such as using complex Army-relevant tasks, examining the impact of trainee characteristics, and assessing various types of transfer performance. The results contribute to a more complete body of knowledge concerning training complex cognitive tasks. The experiments examined research issues related to the effectiveness of different levels of guidance in exploratory learning; the benefit of part-task training for tasks that have sequential and concurrent subtasks; the effectiveness of worked examples for a complex planning task; the relative benefit of constant difficulty, fixed/increasing difficulty, and adaptive difficulty in training; and the usefulness of adaptive practice. These experimental results were used to update the meta-analytic results, which were then used to implement the two additional objectives. Specifically, we developed algorithms to quantify the relationships between the six training methods, performance, and various moderating factors, and then implemented these algorithms into a user-friendly graphical user interface tool, called TARGET (which stands for Training Aide: Research and Guidance for Effective Training).

Findings

   The utilized combination of literature review, meta-analysis, and experimentation enabled us to amass qualitative and quantitative evidence on the effectiveness of six different methods for training complex cognitive skills. This evidence was then used to develop algorithms that quantify the relationships between the training methods, performance, and various moderating factors (e.g., trainee characteristics, the type of task/skill to be trained). These algorithms can be used to perform tradeoff analyses for different combinations of training methods. The algorithms make the research findings from this project available to the Army training, development, and research communities, allowing users to systematically explore training methods that would be effective for acquiring various cognitive skills.

   The research findings and algorithms were used to develop a user-friendly graphical user interface tool, called TARGET. This tool summarizes the cognitive skill training research and identifies the conditions under which a particular training method is more or less effective. TARGET contains several visualization tools, such that in-depth statistical knowledge is not required to benefit from this tool. TARGET is a web-based tool, which is publicly accessible at **http://bldr-webtest.alionscience.com/Target/**.

Utilization and Dissemination of Findings:

   TARGET, as well as its underlying research database and algorithms, are expected to have utility for a variety of different users. Training developers (with varying levels of expertise) can use TARGET's evidence-based recommendations to identify the most effective training method given a set of desired factors. Ultimately, it can help training developers examine the research evidence related to their particular training situation (i.e., the type of trainees, the type of task/skill to be trained) and inform the selection of training methods to satisfy these particular training needs. Training researchers may benefit from TARGET and the underlying research database by better understanding the state of the training literature, including any possible gaps in the field's understanding of effective methods. Military service program managers can also use this information to direct future research to fill identified gaps or investigate currently inconclusive findings. The capabilities represented in TARGET can serve a number of potential future applications as well, such as expanding the research database to include new training methods (e.g., behavior modeling) or task/skill types (e.g., interpersonal skills) and/or adapting the tool's architecture to a different literature domain beyond training. Finally, the developed algorithms can be applied in a variety of applications beyond TARGET, such as serving as input for human performance models to analyze the impact of different training or technological approaches.

UNDERSTANDING THE IMPACT OF TRAINING ON PERFORMANCE

CONTENTS

UNDERSTANDING THE IMPACT OF TRAINING ON PERFORMANCE

Project Overview

Background / Research Need

The overall goal of this four-year program of research was to understand how training methods could be effectively employed to improve performance on Army-relevant tasks. Many factors interact to influence the effectiveness of training. Recent theory and research has resulted in training models that consider the large space of variables that influence training effectiveness (e.g., Alvarez, Salas, & Garofano, 2004; Colquitt, LePine, & Noe, 2000; Tannenbaum, Mathieu, Salas, & Cannon-Bowers, 1991). These influences typically include individual, training, and contextual factors. Individual factors include examples such as trainee experience level, abilities, and motivation; training factors include examples such as instructional methods, delivery mode, and feedback strategies; and contextual factors include examples such as climate for training (Quiñones, 1997) and climate for training transfer (Blume, Ford, Baldwin, & Huang, 2009; see also Alvarez et al., 2004; Tannenbaum et al., 1991). Although each is a burgeoning area of research, significant efforts have been focused on understanding the effects of training factors. This is not surprising, given that training factors are often more malleable than their individual and contextual counterparts, which provides increased control over the training development process (e.g., Baldwin & Ford, 1988).

Extant syntheses of the training literature typically distinguish between two primary training factors that influence training effectiveness: *instructional principles* and *delivery methods* (Alvarez et al., 2004). Instructional principles research examines the effectiveness of different techniques for conveying knowledge or developing skills, such as part-task training (Fontana, Mazzardo, Furtado, & Gallagher, 2009), learner control (Kraiger & Jerden, 2007), behavior modeling (Taylor, Russ-Eft, & Chan, 2005), and error management training (Keith & Frese, 2008). Delivery methods research examines the mode of instruction and information presentation, often comparing traditional face-to-face instructional modes (e.g., lecture) with technology-based modes (e.g., computer-based training; see Arthur, Bennett, Edens, & Bell, 2003; Bayraktar, 2002; Landers, 2009; Means, Toyama, Murphy, Bakia, & Jones, 2009). This distinction is important given that both theory and meta-analytic evidence suggests that learning is impacted more by the instructional principle than delivery method (Clark, 1984, 1994; Sitzmann, Kraiger, Stewart, & Wisher, 2006).

In practice, however, modern training efforts seldom employ a single instructional principle or delivery method. Instead, organizations seek the most effective combinations of instructional principles and delivery methods that satisfy the desired training outcomes. For example, Kozlowski et al.'s (2001) Adaptive Learning System details the process of integrating multiple instructional principles and delivery methods into a single training intervention method targeted at improving learners' adaptation to new or uncertain situations. The strength of such a system relies on first identifying the relevant skills to be trained and then subsequently designing an appropriate training method to develop these skills. These ideas are espoused in the Army Learning Concept 2015 (ALC 2015; U.S. Department of the Army, 2011), the focus of which is to visualize integrated instructional principles and delivery methods suited to develop learner

adaptability. In line with the ALC 2015, our focus moved beyond the instructional principle versus delivery method categorization to instead identify specific *training methods*[1] targeted at improving Army-relevant skills (for details on identifying the relevant methods, see Carolan, McDermott, Hutchins, Wickens, & Belanich, 2011). In particular, we chose to focus on *cognitive skills* (e.g., problem solving, quantitative, spatial reasoning, decision making), as these skills are particularly important for developing an adaptable fighting force (U.S. Department of the Army, 2010). However, the research on training cognitive skills to date lacks the systematic organization needed to provide research-based recommendations for practice.

Accordingly, a central focus of the current research program was to meta-analytically summarize the extant training literature with regards to the effectiveness of different methods for training cognitive skills in complex task domains (e.g., Kalyuga, 2009, 2011; Paas & van Gog, 2009; Sweller, 1988; van Merrienboer, Kester & Paas, 2006). We also focused heavily on identifying moderators of a given training method's effectiveness. That is, we were not only interested in identifying whether a method had an overall benefit (or lack thereof), but in identifying for whom specifically that benefit existed and if this benefit varied according to the training performance outcome(s) of interest and type of cognitive task/skill trained. Further, we examined the benefits of cognitive skill training methods for two important outcomes: (1) acquired knowledge and skills as demonstrated during the training (i.e., learning) and (2) *training transfer* –that is, the degree to which trainees are able to apply and use what they learned once they are in the field or on the job (see Kraiger, Ford, & Salas, 1993). Training transfer is typically defined as the extent to which skills learned in one task context generalize to performance in another task or situation (e.g., Ford & Weissbein, 1997; Wickens, Hollands, Banbury, & Parasuraman, 2013). Transfer can be broken down into near transfer and far transfer, based on the similarity between training tasks and transfer tasks/performance environment. *Near transfer* tasks represent different, yet similar tasks to trained tasks (e.g., Barnett & Ceci, 2002); in contrast, *far transfer* tasks represent tasks much less similar to the training task in terms of difficulty or structure, such that learners must adapt learned skills to these new situations (Ivancic & Hesketh, 2000). Transfer may also be assessed with respect to time, contrasting immediate with delayed transfer. The present research investigated such factors when examining a training method's effect on transfer.

In summary, this research sought to answer the following question: "*Given a certain cognitive task/skill to be trained, a set of trainees, and training outcome(s) of interest – what training methods will be more likely to produce effective learning and transfer*?" The specific objectives used to address this question are explained in further detail below.

Key Research Objectives

As aforementioned, the overarching goal of this research effort was to develop evidence-based guidelines for the relative effectiveness of different training methods for acquiring and transferring cognitive skills in complex task domains. The overall approach for this research effort is shown in Figure 1, and can be broken down into four main research objectives.

---

[1] From this point, we refer to training method as the combination of instructional principles and delivery methods.
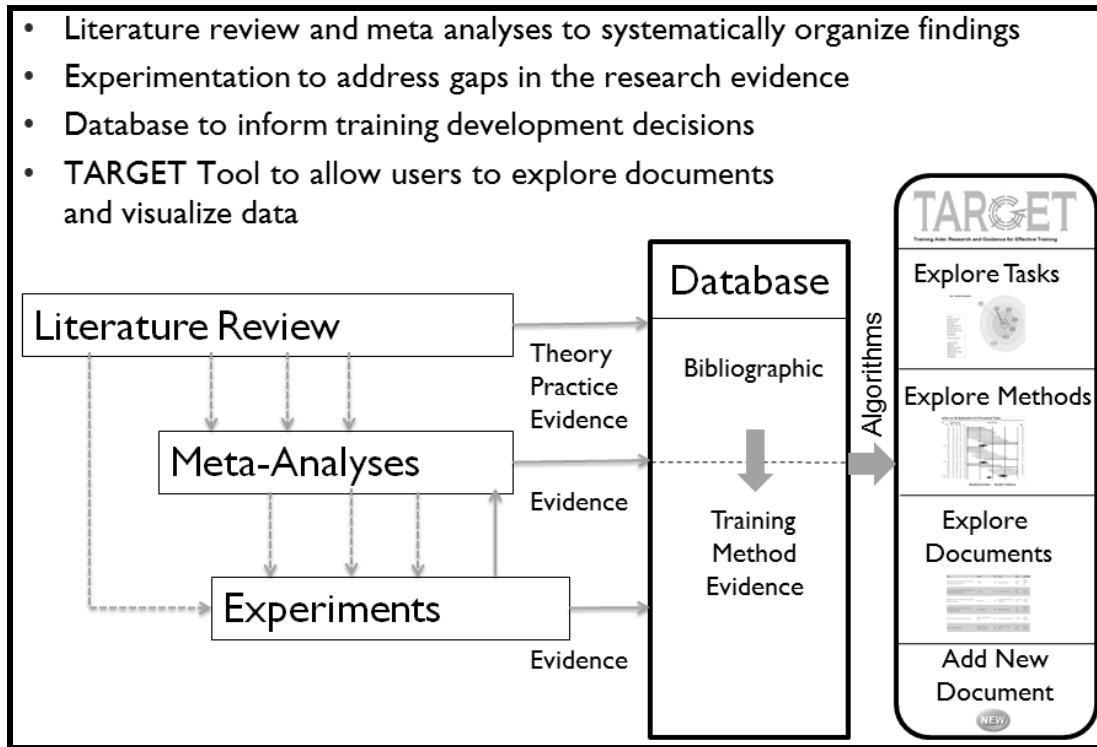
Figure 1. Overall research approach

First, we sought to summarize the current state of the empirical training literature and identify research gaps. As such, we conducted a broad literature search to gather evidence on the effectiveness of various training methods. From the literature review, we narrowed the possible training method options to those identified as most suitable for cognitive skills. Based on this literature review, we conducted six comprehensive meta-analyses. Each meta-analysis summarized the effectiveness of a specific training method and identified key moderators of this effectiveness. The six training methods examined were:

- *Training Wheels.* A training method geared towards reducing the difficulty of the target task during initial learning by reducing training task errors, as well as helping trainees acquire the appropriate schema to assimilate the target task.

- *Scaffolding*. A training method where assistive supports are provided to trainees to ease the demands of task performance. These scaffold supports are incrementally faded out over time until the trainee is executing the whole task independently.

- *Part-Task Training.* A training method that decomposes complex tasks into a series of smaller tasks, each of which is demonstrated and practiced separately before being practiced as a whole task.

- *Increasing Difficulty*. A training method in which parameters of the task are initially set to lower difficulty levels, to reduce the intrinsic load early in training, and then increased as training progresses, until the difficulty reaches the level of the target task. The difficulty

3

levels can increase in either a fixed, pre-determined schedule or adaptively based on the trainee's performance.

- *Exploratory Learning*. A training method in which the trainee explores a task environment on his or her own. The level or type of guidance given to the trainee can vary within this method (e.g., only providing a user manual to reference versus the provision of input by trainers in response to trainee questions).

- *Learner Control*. A training method that provides trainees with decision making control over specific dimensions or activities within a structured learning environment.

Thus, our project provided a broad review of the training effectiveness research, drawing from the training, learning, and military-specific literatures. This review also helped identify inconsistencies and gaps in the literature. Accordingly, our second research objective was to conduct a series of research experiments in order to collect additional research to help fill several identified research gaps from the meta-analyses. A total of five experiments were conducted that included several common design elements. For example, all the experiments involved training complex Army-relevant tasks, such as operating digital systems or planning. Furthermore, learners were required to transfer the training by applying learned material to a new problem, or to a similar problem but using a different technology/ system. The experiments also examined the impact of trainee characteristics (i.e., trainee experience, trainee ability) on the training performance outcomes. By doing so, these focused experiments provided a more complete body of knowledge not only on the overall effects of complex cognitive skill training methods, but also knowledge concerning when, where, how, and for whom are these methods effective.

The third key objective of this research effort was to develop algorithms for identifying the ideal training method given a specific combination of factors. These factors included: performance outcomes (e.g., learning, near transfer, far transfer), task/skill type moderators (e.g., perceptual, psychomotor, cognitive-declarative), and trainee characteristics (e.g., experience, aptitude), among others. These algorithms can be applied in a variety of applications from decision support tools for training developers to input for human performance models to analyze the impact of different training or technological approaches.

Finally, to easily communicate these findings to training developers and researchers, our fourth research objective was to develop a user-friendly graphical user interface tool, called TARGET (which stands for Training Aide: Research and Guidance for Effective Training). This tool summarizes the cognitive skill training research and identifies the conditions under which a particular method is more or less effective. Training designers can use TARGET's evidence-based recommendations to identify the most effective training method given a set of desired factors. In turn, this match between training needs and empirically-supported training methods may improve learning and training transfer. Researchers may also benefit from TARGET by better understanding the state of the training literature, including any possible gaps in the field's understanding of effective methods. Military service program managers can use this information to direct future research to fill identified gaps or investigate currently inconclusive findings.

The remainder of this report discusses in greater detail the purpose, methodology, and results of each of these four research objectives (i.e., literature review and meta-analyses, experiments, algorithms, TARGET tool). We begin by discussing the first stage of this research program – the initial literature review used to gather evidence on the effectiveness of various training methods.

Literature Review

As part of the literature review, we developed a training framework to organize the results gathered (see Carolan et al., 2011). This framework was essential for developing guidelines for training developers by providing the skeleton on which to hang the literature findings. The framework also increased the interpretability of our findings by proving a common language that eliminated much of the jargon associated with different approaches to studying training. The literature review revealed three key consideration/factors which are influential on the effectiveness of training: (1) What needs to be trained (task factors)?, (2) Who needs to be trained (trainee characteristics)?, and (3) What are the performance outcomes of interest (outcome criteria)? The literature review suggested that the relative effectiveness of many training methods are moderated to a greater or lesser degree by interactions among task factors, trainee characteristics, and outcome criteria. In addition to these three factors, practical factors such as cost and availability may limit the training methods available for consideration. Below we briefly summarize some of the key evidence gathered during the literature review stage of this program of research.

Task Factors

Task factors include characteristics such as task/skill type and task difficulty. Task/skill type characterizes tasks by the types of knowledge and skills (e.g., psychomotor skills) required for effective performance. For complex cognitive tasks, the importance of cognitive task analysis as part of the training needs analysis has been advocated by numerous researchers (e.g., Frederiksen & White, 1989; Goettl & Shute, 1996), suggesting that different cognitive skills respond better to different training methods (e.g., Seamster, Redding, & Kaempf, 1997). However, our review of the training literature found few studies that explicitly compared the effectiveness of training methods for different types of cognitive skills. Although there is some evidence that different *delivery* methods (e.g., face-to-face instructional modes, technology-based modes) are more or less effective for psychomotor, knowledge-based, and interpersonal skills (Arthur et al., 2003), prior research has not examined training *methods* as currently conceptualized and has not examined cognitive skills. As such, the key objective of the current research effort was investigating training methods targeted to train complex cognitive tasks in order to identify if different methods were more or less effective for different types of cognitive skills (e.g., problem solving, quantitative, spatial reasoning, decision making).

Regarding task difficulty, a growing body of research suggests that task difficulty influences the effectiveness of various training methods. Task characteristics such as the number of elements/components or information sources, the interactivity of those task components, the degree of task structure, the potential number of solution paths and possible solutions (e.g., Paas & van Gog, 2009) and dynamic changes in component interactivity over time (Wood, 1986) contribute to task difficulty. From the perspective of cognitive load theory (e.g., Paas & van Gog,

2009; Sweller, 1988; van Merrienboer et al., 2006), these intrinsic task characteristics require more cognitive effort on the part of the learner. Two findings related to task difficulty that have been consistently demonstrated in the extant literature are as follows: (1) the organization of the component skills and their interactions have implications for training effectiveness (e.g., Goettl & Shute, 1996), and (2) training methods that are effective for promoting retention and transfer for simple tasks are not always effective for complex tasks (e.g., van Merrienboer et al., 2006; Wood, 1986).

Trainee Characteristics

Trainee characteristics such as aptitude, specific abilities, and experience moderate the effectiveness of various training methods and can be important factors in selecting the optimal training approach (e.g., Gully & Chen, 2010; Snow, 1989). Consistent with cognitive load theory, training methods that reduce difficulty may negatively impact experienced trainees (Paas & van Gog, 2009; van Merrienboer et al., 2006), and may possibly have a negative impact on trainees with high cognitive ability. There is also a body of research on aptitude by treatment interactions that indicate the effectiveness of training interventions can differ depending on trainee aptitude for self-regulation during learning. For example, lower ability trainees tend to benefit from structured lessons and higher ability trainees benefit from less structured training (Bell & Kozlowski, 2008; Snow, 1989).

Outcome Criteria

Desired training effectiveness outcomes typically involve knowledge acquisition and retention (commonly referred to as learning), and transfer of knowledge/skills learned during training to performance in an operational environment. Accordingly, the effectiveness of a training method can be assessed according to a variety of training evaluation criteria (see Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997; Arthur et al., 2003). Two considerations are particularly relevant to understanding training effectiveness. The first is that a given training method may have different consequences depending on whether the performance outcome of interest is learning or transfer. Some highly structured training methods that aid skill acquisition (i.e., a learning criterion) have been found to be less effective or even have a negative effect on transfer (Healy & Bourne, 2011; Schmidt & Bjork, 1992). In addition, there is evidence that some less structured training methods that require more learner effort can yield higher performance during transfer despite lower in-training learning performance, especially when skills are transferred to new problem situations (e.g., Schmidt & Bjork, 1992; van Merrienboer et al., 2006). This has been referred to as the *transfer crossover effect* (Bell & Kozlowski, 2008) or *transfer paradox* (van Merrienboer et al., 2006).

A second consideration is that differences in the definition and measurement of transfer may have implications for evaluating training effectiveness. Transfer is defined in terms of the extent to which knowledge and skills learned in one context influence performance in another context (Wickens, Hollands, Banbury & Parasuraman, 2013). Researchers have characterized the similarity between the training task and transfer task as the 'near-far transfer distinction' (e.g., Barnett & Ceci, 2002). *Near transfer* involves application of skills to a task or situation very similar to the training task. In contrast, *far transfer* involves application of skills to a task or

situation *different* than the training task. The transfer task can be 'far' along a number of dimensions, including transfer to a different performance environment, a more complex task, or an entirely new situation or problem (e.g., Barnett & Ceci, 2002; Keith & Frese, 2005). Correspondingly, *transfer distance* can be defined in terms of the time between when the training task is learned and transfer performance is assessed (i.e., *immediate transfer* versus *delayed transfer*); delayed transfer requires a deeper understanding of the learning content. Transfer of training is therefore a broadly defined concept that requires understanding which type(s) of transfer are of interest to inform training design.

Summary of Six Meta-Analyses

The meta-analyses were a key tool in synthesizing research findings and identifying research gaps in the extant training literature. The benefit of a meta-analysis is that it can empirically summarize the collective wisdom on a topic. It also provides a way to systematically evaluate the impact of specific moderators on given relationships of interest, such as the impact of trainee experience on the relationship between the part-task training and transfer. In other words, meta-analysis does not just provide an overall rating of whether part-task training "benefits" or "costs/hinders" transfer, but rather provides insight into the specific conditions (e.g., trainee experience, trainee ability, transfer distance, task/skill type) under which this training method may amplify or diminish its influence on transfer performance. Accordingly, by conducting moderator analyses within the meta-analyses, we had the capability to quantify the relationships between training methods and performance outcomes under various conditions. Note that these moderator findings became the basis for the algorithms and weightings that underlie the TARGET tool.

Methodology

In terms of the methodology, both a transfer ratio (TR) and Hedges' *g* were used in the meta-analyses. The TR is a ratio of the *treatment* group's performance to the *control* group's performance. Treatment refers to the experimental group who receives the training method, while control refers to the group in the research study who receives no training or a lesser degree of the given training method. Ratios similar to the TR have been employed in other meta-analyses to express degree of benefit, such as a ratio of the advantages for multi-modal over single mode displays (Lu, Wickens, Hutchins, Sarter, & Sebok, 2013). Similar to the more commonly used odds ratio, TRs less than 1 indicate a 'cost' for the training treatment, numbers greater than 1 indicate a 'benefit,' and a value of 1 indicates 'no difference' between treatment and control. The usefulness of the TR method is that it expresses a benefit (or cost) in a way that is directly interpretable to the user (Lu et al., 2013; Wickens, Hutchins, Carolan, & Cumming, 2013). For example, a ratio of 1.3 means "30% more effective."

The TR was complemented with the Hedges' *g*, an effect size metric for comparing treatment and control group standardized mean differences (Rosenthal, 1991). The *effect size* is a statistical concept that measures the strength of the relationship between two variables (Preacher & Kelly, 2012). Interpreting Hedges' *g* focuses on the deviation from 0 with negative values indicating a cost for the training treatment, positive values indicating a benefit and a value of 0 indicating no difference. The value of Hedges' *g* is that it is a more conventional metric with a structured way to characterize statistical power. A key methodological contribution of the current

research project is that across our meta-analyses in which both were deployed, the TR and Hedges' *g* provided convergent information. This provides justification for the use of the easily-interpretable TR metric, especially in cases where the data required to calculate Hedges' *g* is not available for many of the primary studies, which would decrease the statistical power of Hedges'*g*. All the findings (TR and Hedges' g) refer specifically to transfer performance.

In terms of study inclusion criteria for the six meta-analyses, for an identified study to be included, the participants could not be school-aged children or the elderly to ensure generalizability to typical Army trainees. We also required that the study included a control group, and that performance measures gathered from the control group mirrored those received from the treatment group.

The choice of moderator variables was largely driven by cognitive load theory (Paas, Renkl, & Sweller, 2003); however, we also examined variables known to be important for Army training, as well as others not directly covered by cognitive load theory. Cognitive load theory posits three different types of resources demands: intrinsic load, germane load, and extraneous load. Demands associated with task difficulty are called *intrinsic load*. Demands associated with learning and skill acquisition are called *germane load*. Demands that are neither intrinsic nor germane are called *extraneous load*. Examples of extraneous load include distractions in the environment or a poorly designed learner interface. Cognitive load theory predicts that lessening the intrinsic and extraneous loads will free more resources for germane load and thus increase learning and transfer. Therefore, reducing task difficulty, at least in the early stages of a training program, should increase learning and transfer performance. In the same manner, for the less experienced learner, the task will be more complex and hence that learner will benefit more from a load reducing training method (Rey & Buchwald, 2011). As such, we coded for moderator variables such as task difficulty and trainee experience.

In addition to the cognitive load theory-relevant moderators, our review identified a number of other moderators commonly examined in the literature, such as instructor presence, as well as training method-specific moderators. By training method-specific moderators, we mean training design features that varied *within* a given training method. For example, for the training method part-task training, a training method-specific moderator was whether the different component tasks were trained concurrently or sequentially. As another example, for the scaffolding training method, one important moderator was whether the scaffolds were removed according to the trainee's progress/performance during the training (i.e., adaptive scaffolding) or on a set schedule regardless of trainee performance (i.e., fixed scaffolding).

Finally, although our focus was on training methods facilitating cognitive load reduction, we note that two of the training methods examined in this research, exploratory learning and learner control, are not designed to reduce cognitive load. Rather, their effectiveness is derived from the different theoretical perspective, suggesting certain *advantages* induced by active learning and learner engagement (e.g., Bell & Kozlowski, 2003; Dunloski, Rawson, Marsh, Nathan, & Willingham, 2013; Roediger & Karpicke, 2006). However, it is important to note that some research has found such engagement may also undesirably increase cognitive load (discussed further below). Key meta-analytic findings from the six examined training methods are described next.

Error Prevention: Training Wheels and Scaffolding

The training methods of training wheels and scaffolding both seek to prevent errors, especially in the early phases of training. *Training wheels* are typically operationalized in one of two ways: lockouts and worked examples. In lockouts, certain features of a computer system are unavailable and then are incrementally made available as trainee progresses through the training. In worked examples, trainees are given complete or partial worked-out solutions in order to learn correct and efficient strategies. For the *scaffolding* training method, assistance is provided to trainees in the early phases of instruction. This assistance can help focus the trainees' attention or simplify the task. While prior literature reviews have been conducted for both training wheels (Carroll, 1990; Shen & Tsai, 2009; van Gog, Paas, & Sweller, 2010; van Gog & Rummel, 2010) and scaffolding (Pea, 2004), to the authors' knowledge, no meta-analyses have been conducted on these two training methods. Accordingly, we conducted meta-analyses on these methods to examine their training effectiveness under different conditions. Some key hypotheses and findings from these meta-analyses are highlighted below. For complete details, please see Hutchins, Wickens, Carolan, and Cumming (2013).

Hypotheses. From the perspective of cognitive load theory (van Gog et al., 2010), error prevention training methods were hypothesized to reduce intrinsic load and extraneous load early in skill acquisition and thus support learning by availing more resources for germane load. We predicted that inexperienced trainees (i.e., being unfamiliar with the task) would experience high intrinsic load and thus would benefit from error prevention strategies more than experienced trainees (e.g., Rey & Buchwald, 2011). Finally, we were uncertain of the overall benefit of the two training methods given the following inherent tradeoff. On the one hand, reducing intrinsic load should favor these two training methods; however, on the other hand, partially preventing full choice of learner options may inhibit full engagement in, or "active learning" of, the task, which could itself inhibit learning and transfer (Keith & Frese, 2008; Kraiger & Jerden, 2007; Roediger & Karpicke, 2006). As such, we did not predict in advance the relative weighting of these two counteracting influences on skill acquisition and transfer.

Findings. Thirty-one studies identified in the extant training literature met our inclusion criteria for the training wheels meta-analysis, yielding 74 Hedges' *g* estimates and 79 TR estimates. In this, and subsequent meta-analyses, we treated multiple effects within a study and across studies as equally independent effects. Overall, the results showed a 30% transfer benefit for training wheels compared to unsupported or less supported training (TR = 1.3, *g* = +0.21). This benefit was moderated by trainee experience, instructor presence, and transfer type. As predicted, non-experienced trainees benefited more from training wheels (not enough studies to calculate TR, *g* = +0.44) than experienced trainees (not enough studies to calculate TR, *g* = +0.28). Interestingly, the presence of an instructor mitigated any benefits of training wheels (TR = 1.18, *g* = +0.14); trainees performed better when the instructor was absent (TR = 1.51, *g* = +0.37). Finally, the further the transfer, the less the benefit of training wheels; that is, training wheels

benefited near transfer (TR = 1.56, g = +0.34), but did not benefit far transfer to a more complex task (TR = 0.91, g = -0.08). In terms of the training method-specific moderators, a benefit was observed when the training wheels were implemented as worked examples, a training strategy that provides learners with a worked out solution or partially completed steps towards the solution in order to prevent the use of weak, inappropriate, or inefficient strategies (van Gog & Rummel, 2010; TR = 1.29, g = +0.31). A benefit of lockouts was observed (TR = 1.1, g = +0.30).

Only eight scaffolding studies identified in the extant training literature met our inclusion criteria, yielding 21 Hedges' g estimates and 23 TR estimates. Overall, the meta-analytic results suggested a large, 60% benefit for scaffolding (TR = 1.58, g = +0.46). This effect was moderated by trainee experience, instructor presence, transfer distance, and the schedule for removing scaffolds. Contrary to cognitive load theory predictions and the results of the training wheels analysis, trainees with experience showed a larger benefit of scaffolding (TR = 1.70, g = +1.34) than those without experience (TR = 1.30, g = +0.26). Transfer was better when the instructor was absent (TR = 1.81, g = +1.09) than when the instructor was present during training (TR = 1.52, g = +0.27). In terms of a training method-specific moderator for scaffolding, the scaffolding aids could be removed using a fixed schedule or could be removed adaptively in response to trainee performance. We found there was a larger benefit to removing the aids adaptively (TR = 1.63, g = +0.72) than on a fixed schedule (TR = 1.56, g = +0.29). The benefit of scaffolding to near transfer tasks (TR = 2.03, g = 0.66) was stronger than the benefit to identical transfer tasks (TR = 1.55, g = 0.44). There were no studies examining scaffolding and far transfer.

In summary, the meta-analytic findings support the use of both these error prevention methods as a way to reduce intrinsic and extraneous loads during training. Consistent with cognitive load theory, training developers should consider creating an adaptive strategy for removing error prevention mechanisms in response to trainee performance. This adaptive removal has the potential to benefit transfer performance. Lastly, training developers should also take into account the trainee experience when designing a training program, reducing the stringency or aggressiveness of error prevention for experienced trainees.

Part-Task Training and Increasing Difficulty

Hypotheses. Both part-task training and increasing difficulty training methods manage difficulty in the early phases of training. Part-task training divides tasks into more manageable subtasks, and increasing difficulty simplifies early tasks gradually shifting to more difficult tasks. Cognitive load theory predicts that this reduction of intrinsic load should free more resources for learning and thus benefit performance outcomes. Yet, such early-in-training decomposition or simplification of the task may trigger unintended negative consequences that offset any training benefits. Indeed, the extant literature shows mixed success for these two methods. In a review of psychomotor tasks, Wightman and Lintern (1985) found that part-task training was successful, but only if the sub-tasks were performed sequentially (not concurrently) in the full task. A potential drawback is that part-task training does not offer an opportunity to practice timesharing skills; for example, the ability to scan for targets *while* operating a vehicle (Damos & Wickens, 1980; Lintern & Wickens, 1991). Note that one prior meta-analysis was located for part-task training (Fontana et al., 2009). This prior meta-analysis focused solely on psychomotor skills;

whereas our meta-analysis emphasized cognitive skills and therefore included primarily studies not included in the Fontana et al meta-analysis.

For the increasing difficulty training method, task parameters may initially be set to low difficulty and then increase throughout training. This increase can occur on a fixed/increasing schedule (i.e., difficulty level increases across training on a set schedule) or it can be adaptive (i.e., changes in difficulty level across training based on trainee performance). Wightman and Lintern (1985) did not find a benefit of increasing difficulty training, but they did not systematically contrast adaptive and fixed/increasing schedules (Mane, Adams, & Donchin, 1989; Metzler-Baddeley & Baddeley, 2009). The drawback for increasing difficulty method is that by presenting a simplified task, the trainee may learn an inappropriate version of the task that is dissimilar to the full-difficulty transfer task.

We conducted meta-analyses on part-task training and increasing difficulty to better understand under which conditions these methods should be used and in which conditions they should be avoided in order to better clarify the mixed results in the literature. As with error prevention methods, we were unsure how much the drawbacks might offset the cognitive load theory-based benefits. Nevertheless, consistent with cognitive load theory, we did hypothesize greater benefits for non-experienced trainees and fewer benefits for simpler tasks. For part-task training, we predicted greater benefits for sequential rather than concurrent tasks. Some key findings from these meta-analyses are highlighted below. For complete details, please see Wickens et al. (2013).

Findings. Twenty-two studies identified in the extant training literature met our inclusion criteria for part-task training yielding 65 contrasts or effects between a part-task and whole task condition (65 TR estimates and 35 Hedges' g estimates). Overall, there was a 13% cost for part-task training (TR = 0.87, g = -0.06). A key finding was related to the timing of subtasks. There was neither a cost nor benefit of part-task training if the subtasks were performed sequentially in transfer; in contrast, there was a strong cost to part-task training if the subtasks were *concurrent* in the transfer tests (TR = 0.71, g = -0.35). This supports the assertion that part-task training does not provide the necessary opportunity to practice timesharing skills. The importance of practicing timesharing skills is indirectly supported by the substantial benefit found for 'variable priority training' in comparison to fixed difficulty studies (TR = 1.27, g = +0.74). Such *variable priority training*, in which the whole task is maintained but different aspects of the task are emphasized or de-emphasized, combines the best of both worlds because the whole task remains intact.

The impact of part-task training was also found to be influenced by several moderator variables. The part-task training costs were moderated by task difficulty, with more difficulty tasks showing smaller costs (TR = 0.83, g = -0.10) compared to less difficult tasks (TR = 0.60, g = -0.49). Correspondingly, there was some (although modest) evidence that experienced trainees suffered more from part-task training (and benefitted more from whole-task training). (Experienced: TR = 0.84, g = -0.93; Novice: TR = 0.84, g = -0.25) Both of these effects are consistent with predictions of cognitive load theory. When an instructor was present in the delivery environment there was a cost to part-task training (TR = 0.85, g = -0.43), but not when the instructor was absent (TR = 1.11, g = +0.40).

A search of the increasing difficulty training literature identified yielded 15 Hedges' *g* estimates and 30 TR estimates that used fixed/increasing difficulty training as the control/comparison condition). The increasing difficulty meta-analysis combined simplification part-task training studies and the adaptive training literature under the larger umbrella of increasing difficulty. Overall, there was neither a significant cost nor a benefit to the increased difficulty training method (TR = 1.22, *g* = +0.03). Transfer effects were moderated by trainee experience, instructor presence, transfer distance, and type of transfer test. Consistent with cognitive load theory and the meta-analyses reported above, non- experienced trainees benefitted from the increasing difficulty training method (TR = 1.10, *g* = +0.75); there were no data for experienced trainees. As above, the load-reducing increasing difficulty training method benefitted performance when the instructor was absent (TR = 1.38, *g* = + 0.42), but not when present (TR = 0.89, *g* = -0.48). Benefits were also observed when transfer was immediate (TR = 1.39, *g* = +0.46) versus delayed (TR = 0.95, *g* = -0.36), and when transfer difficulty was similar to training difficulty (TR = 1.32, *g* = +0.21) versus near transfer (TR = 0.70, *g* = -0.56). The results regarding the impact of the task difficulty moderator were inconclusive due to few primary study data points.

The adaptive nature of increasing difficulty also had a significant effect. When such an increase was adaptive and based on the trainee's performance, the relative benefit compared to fixed/increasing difficulty was substantial (36%). However, increasing difficulty on a fixed/increasing schedule without considering trainee performance produced a significant (23%) transfer cost. Thus, adaptive difficulty schedules produce the benefits expected by cognitive load theory (Sweller, 2010).

One implication of these findings is that training developers should adopt methods that concurrently embed timesharing tasks within whole task environment. An example of one such task environment is a Soldier who must consult a navigational device while on the move. A clear alternative to part-task training methods that also enables practice of timesharing skills is the variable priority method described above. Another implication is that fixed/increasing difficulty schedule may not be the most effective approach for all trainees. Instead, adapting the difficulty to trainee performance has a higher potential to increase transfer performance. Furthermore, our moderator results suggest that trainers should consider trainee characteristics when choosing a specific method given that both part-task training and increasing difficulty were more beneficial for non-experienced trainees.

Learner Control and Exploratory Learning

Hypotheses. Active learning methods such as learner control and exploratory learning are designed to promote trainee engagement in the learning content and process (Bell & Kozlowski, 2008; Means et al., 2009). Learner control does this by giving trainees control over at least one aspect of training such as instructional pace (Orvis, Fisher, & Wasserman, 2009), content sequencing (Tang, 2004), how much of the lesson to review (Taylor, 2005), the amount of practice to engage in (Schnackenberg & Sullivan, 2000), the amount of feedback given (Pridemore & Klein, 1991), or whether to receive system-generated advice for the next task (Shyu, 1993). Learner control has been found to impact positively learning (e.g., Orvis, Fisher, & Wasserman, 2009).

Exploratory learning promotes engagement by letting trainees explore learning content to discover key relationships and interactions on their own. Exploratory learning can vary in the amount and type of guidance provided (Bell & Kozlowski, 2008). Proponents of exploratory learning argue that this method encourages metacognitive activity and self-regulation of learning; which, in turn, can aid the development of adaptable and complex skills (e.g., Bell & Kozlowski, 2008; Heimbeck, Frese, Sonnentag, & Keith, 2003; Keith & Frese, 2005).

Despite the promise of both learner control and exploratory learning, the literature has shown mixed evidence for their effectiveness (e.g., Clark, 2009; Corbalan et al., 2011; Doolittle, 2010; Smith, Ford, & Kozlowski, 1997; Kirschner, Sweller, & Clark, 2006; Mayer 2004). For example, learner control has been found to benefit skill-based transfer tests (Mayer & Chandler, 2001; Doolittle, 2010), yet mixed effectiveness for knowledge acquisition (e.g., Doolittle, 2010; Orvis et al., 2009). Exploratory learning has been found to benefit tasks that are primarily procedural (e.g., Carroll, 1990), but not problem-solving tasks (e.g., McDaniel & Schlager, 1990).

Accordingly, we conducted meta-analyses of learner control and exploratory learning to identify moderating factors that may account for the mixed effectiveness observed for these training methods (Carolan, Wickens, Hutchins, & Cumming, in press). Our meta-analysis argued that learner control and exploratory learning enable more "freedom" to control and explore, respectively. We were interested in both the costs and benefits of such freedom during training. Hence, as described below, we considered program control (i.e., low learner control) and restriction of exploration to be manifestations of *reduced* cognitive load. We highlight some key findings from these meta-analyses below; for full details, please see Carolan et al. (in press).

As above, we hypothesized that the reduced cognitive load of less freedom might offset the costs of less active learning; and as such, the net effect of these two influences would be difficult to predict *a priori*. But also as above, we hypothesized greater benefits (or reduced costs) of greater freedom for experienced trainees. We also predicted greater benefits (or reduced costs) for far transfer and delayed transfer tasks versus near and immediate transfer tasks; thereby producing the transfer crossover effect. In these two meta-analyses, only Hedge's g was calculated due to contractual time constraints.

Findings. Forty studies from the extant training literature that met our inclusion criteria for learner control training yielded 144 Hedges' g estimates between a more and a less controlled learning environment. The overall analysis indicated no overall average cost or benefit from using learner control as a training method ($g = +0.02$). This null effect, however, was moderated by task/skill type, trainee experience, and type of learner control provided to the trainee. In terms of the moderator of task/skill type, learner control produced a small cost for factual knowledge ($g = -0.06$), but benefitted both procedural and problem solving tasks ($g = +0.09$). The effects were also moderated by trainee characteristics in that learner control benefited trainees with prior experience ($g = +0.43$), but not novices ($g = -0.004$). In terms of the 'type of learner control' moderator, there was a significant benefit when trainees could control their pace ($g = +0.15$), but a significant cost to transfer performance when trainees could control the amount of feedback or practice ($g = -0.19$). We observed no effect of transfer distance as a moderator for the learner control-performance relationships.

Studies of exploratory learning training from the extant training literature yielded 135 effects. The overall analysis indicated a small but significant overall benefit for more guided exploration ($g = +0.15$), that is, when learners are encouraged to explore certain screens of a computerized instruction program. This was moderated by task/skill type, transfer test type, transfer distance, and trainee experience. In terms of the task/skill type moderator, whereas more exploration produced a cost for factual knowledge ($g = -0.57$) and problem solving tasks ($g = -0.28$), this method benefitted procedural tasks ($g = +0.11$). Exploratory learning resulted in a cost for near transfer ($g = -0.13$) but benefited far transfer ($g = +0.16$), and the benefit of exploratory learning increased as transfer distance increased. Similar to the learner control meta-analysis, the exploratory learning meta-analysis showed a benefit to trainees with prior experience ($g = +0.34$) but not novices ($g = -1.06$).

In summary, learner control and exploratory learning are both effective under certain conditions that overlap considerably. Both training methods are more effective for learning to perform a cognitive skill than for recalling factual knowledge. Both training methods benefitted trainees with prior experience but not novices. In contrast, learner control had the most benefit to very near transfer, while exploratory learning had the most benefit to far transfer, suggesting that learner control and exploratory learning may operate by different mechanisms for influencing transfer performance. For exploratory learning, the emphasis on finding procedures and strategies for generating rules and solutions may benefit far, adaptive transfer (e.g., McDaniel and Schlager, 1990). For learner control, on the other hand, having greater control freedom may allow the learner to manage cognitive load during training; thereby, increasing the opportunity for learning, which is realized in very near transfer.

Insights from Across the Six Meta-Analyses

Across six meta-analyses, we examined three ways in which training methods were reflected within the context of cognitive load theory.

- In Wickens et al. (2013), we examined two training methods designed to **explicitly** reduce intrinsic load by either subdividing the task into parts (part-task training) and then later reassembling them, or by adjusting parameters of a whole task to lesser workload levels early in training and then increasing difficulty as training progressed (increasing difficulty).

- In Hutchins et al. (2013), we examined two methods designed to **implicitly** reduce cognitive load by discouraging or preventing errors. These methods had two different effects on workload: (1) by locking out certain error-likely options, or guiding correct option selection, the *intrinsic task load* on learner choice was reduced, and (2) by preventing certain "catastrophic" errors in the learning progress (e.g., deleting a file while learning text editing, or crashing a flight simulator) this could eliminate the *extraneous load* of error recovery.

- In Carolan et al. (in press), the primary focus was not on reducing cognitive load, but on providing greater learner **freedom** to control the learning process and explore. Hence, the primary focus was on the presumed benefits of active learning. In contrast, where such

freedoms were restricted (i.e., low learner control and limited exploratory learning), there may also be a benefit from a possible reduced load effect resulting from choice constraints. This reduced load corresponds to the effect predicted in the implicit manipulation of load via error prevention as described above. In other words, it is possible that restricting learner control/exploratory learning has effects similar to preventing errors.

Table 1 provides a summary of the major results of the six meta-analyses. We grouped these results according to the three methods described above for managing cognitive load – Explicitly, Implicitly, and Restricting Freedom. For each column, the control group represents the condition associated with an **increase** in cognitive load and the treatment group represents a **decrease** in cognitive load. For example, "No Learner Control" and "No Exploratory Learning" are the two treatments associated with a cognitive load reduction for the Restricting Freedom meta-analyses. A positive *g* or TR greater than 1.0 indicates that reducing cognitive load improved training effectiveness; a negative *g* or TR less than 1.0 indicates that reducing cognitive load decreased training effectiveness (i.e., was a cost).

Table 1.
Summary of Major Results from the Meta-Analyses

Cognitive Load	Explicitly Manipulated		Implicitly Manipulated		Implications of Restricting Freedom	
Effects	*Part-task Training*	*Increasing Difficulty*	*Training Wheels*	*Scaffolding*	*No Learner Control*	*No Exp. Learning*
1. TR	0.87*	1.22*	1.30*	1.58*	--	--
2. Hedges' *g*	-0.06	+0.03	+0.21*	+0.46	0	+0.15*
3. Trainee Experience Effect?	Yes	Yes	Yes	No	Yes	Yes
4. Task Difficulty Effect?	Yes	No Effect	--	--	--	--
5. Transfer Distance Effect?	No effect	Yes	Yes	No	Yes	Yes
6. Cost of Instructor Presence?	Yes	Yes	Yes	Yes	--	--
7. Adaptive Schedule?	--	Helps		Helps	--	--

Note. * indicates p< .05, -- indicates insufficient data for a comparison to be made.

Overall training effectiveness. The first and second rows describe the *overall* training effectiveness as assessed by the TR or Hedges' *g*. There was a great degree of consistency across the two measures. In all cases, the effect signs were consistent even if statistical significance was not. Note that TRs were not computed for the Restricting Freedom meta-analyses (i.e., for learner

control and exploratory learning). Comparing columns, we note a range of effects. Reducing cognitive load through part-task training is clearly **not** effective overall as indicated by a negative Hedge's *g* and a TR less than 1.0. Reducing load implicitly through error reduction strategies (i.e., via both training wheels and scaffolding) was rather effective. Reducing load by increasing difficulty and restricting exploratory learning (i.e., no exploratory learning) produced a modest benefit to transfer. However, it is important to stress that the effectiveness of all of training methods was significantly dependent on several moderator variables, which are listed in Rows 3-7. Note that these rows indicate only that a moderator effect was present, but do not distinguish whether this effect was identified by TR, Hedges g, or both. The overall trends concerning these moderators are summarized below.

Trainee experience and task difficulty. Both the trainee *experience* effect and the task *difficulty* effect are closely related within cognitive load theory (see Rows 3 and 4 in Table 1). Greater trainee experience and less complex tasks are both assumed to reduce the need for resources allocated to germane load, and hence reduce the benefit of these training methods. In some cases, this reduced benefit (resulting from the reduced need for resource reduction by the experts) may be augmented by an increased drawback of the cognitive load theory-based method (e.g., such as the increased dependency or lack of active learning that training wheels fosters). Thus, either a reduced benefit or an actual increased cost is still consistent with predictions of cognitive load theory, when it is coupled with these effects, such as reducing active learning. Of the eight meta-analytic comparisons conducted for these two moderator variables, six showed significant effects consistent with this cognitive load theory prediction. That is, these methods for reducing cognitive load had a smaller benefit, or possibly even a cost, on effectiveness as trainees become more experienced or the task becomes less complex. One method showed neither a cost nor a benefit, and one showed the opposite effect. We note that the scaffolding meta-analysis was populated by the smallest number of comparisons ($k = 8$, 23 effects), which may explain why this method showed the opposite effect.

Transfer distance. The fifth row presents the commonality of findings across the *transfer distance* moderator, collapsing results from both temporal distance (i.e., immediate versus delayed transfer) and similarity distance (i.e., near versus far transfer). A distance transfer moderator effect indicates that performance improved as distance increased. We assumed that training methods reducing cognitive load might lead to shallower learning as they inhibit deeper engagement/processing. This may be direct, as in the case of removing the freedoms to control and explore (the restricting freedom methods), or indirect by preventing errors (the implicitly manipulated methods). In other words, cognitive load-reducing methods that may improve immediate learning may also hinder transfer of the learned material. The data provide some support for this hypothesis. Four of the meta-analyses provided evidence that reductions of cognitive load had the hypothesized effects across transfer distance. As with the experience effect, the scaffolding method appears to show a trend that contradicted those of the other methods.

Presence of an instructor. The sixth row presents one unexpected effect. That is, the presence of an instructor during training produced a cost to training effectiveness in four of the meta-analyses. The origins of this effect are unclear. To the extent that the instructor might be an added source of extraneous load, this influence should be more pronounced in the control conditions (with higher cognitive load) and hence work in the **opposite** direction. We might

speculate that the effect instead is due to interference or inconsistent feedback between the instructor and the computer-based training system. Input and feedback from the instructor and the computer technology or pedagogical tool underlying the strategy needs to be consistent. Otherwise, the instructor may inadvertently mitigate the effectiveness of the training method.

  Fixed versus adaptive schedule. The final row describes the two methods in which fixed/increasing versus adaptive schedules were implemented. The fixed/increasing schedule was the same for all trainees and independent of trainee performance, whereas the adaptive schedule was modified across the training based on trainee performance. Here again, the findings were consistent. Transfer effectiveness was improved through adaptive strategies whereby difficulty is increased (or scaffolding is removed) contingent upon the individual trainee's performance/degree of skill mastery. Such a finding, while not surprising, does have other training implications, as adaptable schedules become somewhat more complex/costly to implement than the simple "one size fits all" approach.

Overview of Experiments and Additional Literature Review

  We conducted five experiments in order to help address key research gaps identified in the meta-analyses. We also conducted a focused literature review on interpersonal skills to examine its viability as a future research direction beyond the present effort's focus on complex cognitive skills. In the following sections, the rationale behind each experiment, as well as its general methods, key findings and implications are described. When an effect is described as "less" or "more" this indicates significantly less or more. Following the discussion of the experiments, the interpersonal skills literature review is briefly summarized.

  Before discussing the experimental results, it is important to keep a few points in mind. First, the purpose of conducting these experiments was to add their findings to the body of research included in the meta-analyses; as such, these findings have already been incorporated into the meta-analytic results presented in the prior section. Secondly, one should not be particularly focused on whether or not these findings align with meta-analytic results. As aforementioned, the key advantage of meta-analysis (vs. primary experiments) is that meta-analysis can empirically summarize the collective wisdom on a topic. As such, while we argue that these five experiments add meaningfully to the training literature (particularly given their emphasis on Army tasks), they reflect only a relatively small number of data points compared to the more comprehensive meta-analytic results.

Exploratory Learning Experiment

  There were several motivating factors for performing additional research on Exploratory Learning. First, the literature review and meta-analysis showed that exploratory learning has a history of benefiting both near transfer (e.g., Kamouri, Kamouri & Smith, 1986) and far transfer (e.g., Barnett & Ceci, 2002; Keith & Frese, 2005). The active learning aspect of exploratory learning provides opportunities for reflection and deep processing required for learning from errors (Frese, Bodbeck, Heinbokel, Mooser, Schleiffenbaum, & Thiemann, 1991; Keith & Frese, 2008). Second, consideration of cognitive load theory highlighted the fact that the increased effort in searching the problem space for the right rule or strategy to apply, especially in a complex

domain, could cause frustration. This frustration is a source of extraneous load and can inhibit learning. Cognitive load theory then predicts that experienced trainees would benefit more from exploratory learning because their experience mitigates the difficulty (intrinsic load) and frustration (extraneous load). Third, our exploratory learning meta-analysis (Carolan et al., in press) found a benefit of exploratory learning to transfer performance when external guidance was limited and when tasks were primarily procedural. However, the meta-analytic findings were inconclusive regarding which levels and type of guidance (during exploration) were effective. In summary, the extant literature suggests that exploratory learning is useful for transfer. Nevertheless, questions remain regarding how to operationalize this training method and which type of trainees will most benefit.

Experimental methodology. Trainees learned to use a digital interface to interact with simulated unmanned vehicles. Sanders (1999) described basic digital system operator skills as primarily cognitive skills with simple perceptual and motor components involved in performing discrete, multi-step procedures in a digital environment. Trainees used the digital system to specify routes for unmanned air and ground vehicles, edit routes, interact with the digital map, view and classify images, and send reports. This experiment (for additional details, see Carolan, McDermott, Wickens, Fisher, & Gronowski, under review) was designed to compare the effectiveness of learning the task through an exploratory learning process versus a directed training approach without the opportunity for exploration. The experiment looked at the impact of different types of training guidance and trainee characteristics on the relative effectiveness of exploration-based learning.

Trainees experienced one of four training conditions: directed training, guided exploration, minimally-guided exploration, and learner-guided exploration. In *directed training*, the experimenter explicitly described and demonstrated the steps involved in each task. During practice, the trainer provided directive guidance and immediate corrective feedback on all errors. In contrast to directed training, there were three levels of *exploratory training*: guided exploration, minimally-guided exploration, and learner-guided exploration. Instead of receiving directions and demonstrations, trainees in all exploratory learning conditions were encouraged to freely explore the digital system and to consider errors as opportunities for learning. The three exploration conditions increased in the degree of learner freedom as follows. In *guided exploration,* the trainer provided unsolicited coaching as needed throughout practice scenarios when the trainee erred or struggled. Coaching was conceptual to help the trainee think about the problem, but became progressively more explicit if trainees continued to struggle. *In minimally-guided exploration,* the trainee determined the training practice scenario order. The experimenter only provided coaching guidance in response to trainee questions and the conceptual coaching never reached the most explicit level (i.e., suggesting general steps to take). Finally, in *learner-guided exploration,* the trainee determined the training scenario order and there was no trainer to provide coaching or answer questions.

The exploratory learning experiment challenged trainees with a range of transfer tests. Immediately following training, the trainees completed near and far transfer tests in which they had to complete tasks, some new, in the context of an integrated Army mission. Additionally, three weeks after training, the trainees returned to complete far transfer scenarios on a different

digital system. Finally, six weeks after training they returned to complete near and far transfer scenarios on the original digital system.

Findings. Results confirmed the benefit of exploration-based training under certain conditions in a military-relevant task environment with procedural tasks. Consistent with the transfer crossover effect, trainees in the exploratory training conditions (collapsed over three levels of exploratory learning) performed significantly better than those with directed training with regard to far transfer. This was true regardless of trainee experience. The performance on the last session (six weeks after initial training) showed that exploratory training offered a significant retention advantage. The downside for exploratory learning was that the retention and transfer benefit came at a cost of initial training time because it took those trainees longer to complete the actual training.

When the three levels of exploratory learning were examined, the three levels of did not produce any differences in the training performance outcomes. Surprisingly, the learner-guided condition performed as well the other exploratory training conditions. This suggests that learner-guided training without an instructor can be an effective option for learning procedural tasks and offers to reduce training costs associated with fully instructor-guided courseware in the training delivery environment. It may be beneficial to have trainees learn and practice basic procedures on their own (e.g., menu navigation) before starting a training program on the more complex aspects. In terms of trainee characteristics, there was evidence that exploration provided more benefit to higher (vs. lower) ability trainees and those with greater (vs. lesser) prior experience.

Part-Task Training Experiment

Part-task training reduces the full difficulty of a target task during training with the objective of improving the effectiveness and/or efficiency of learning and transfer to the whole target task. Part-task training methods are widely accepted and used in training programs when the full target task is considered too complex or impractical to start training on initially (Lintern, 1991). Part-task training is also used when operational equipment or full mission simulators are not available or deemed too expensive for initial training. By reducing cognitive load, part-task training has the potential to save whole-task learning time, speeding up the learning process and increasing training efficiency relative to whole-task training (Wightman & Lintern, 1985). Yet, Wightman and Lintern's (1985) review concluded that part-task training had limited empirical support overall. The lack of effectiveness of part-task training for concurrent part-tasks (i.e., tasks that need to be performed simultaneously) is thought to be due to the need to also train the timesharing skills between the concurrent tasks, which can only be trained/practiced in the whole task (Goettl & Shute, 1996; Lintern & Wickens, 1991). The results of our part-task training meta-analysis (Wickens et al., 2013) were consistent with Wightman and Lintern's overall findings. However, we found that 'varied emphasis' part-task methods that provided an opportunity to develop timesharing skills are more effective compared to whole-task training. In short, *how* part-task training is implemented influences its effectiveness as a training method.

Results from the literature review and meta-analyses also suggested a number of areas where additional evidence was needed to support training recommendations related to the effectiveness of part-task training. Most of the part-task training research focused on part-tasks

that were segments of a serial task or fractions of an integrated concurrent task, where "integrated" implies a physical interaction or coordination between the two concurrent subtasks (e.g., in flight control or shifting gears). There has been a relatively limited research focused on part-tasks that are concurrent, but separate in the whole task; that is, the subtasks interact only through their concurrent demands on operator attention. This domain includes a variety of task/skill types that involve vehicle control and navigation while monitoring the environment to detect and identify potential threats. We addressed this research gap with a part-task training experiment that consisted of tasks that were concurrent but separate.

Experimental methodology. Our experiment used both a simulation environment and a live environment to train and test reconnaissance using an unmanned ground vehicle (for additional details, see McDermott, Carolan, Fisher, Gronowski, & Gacy, 2013; McDermott, Carolan, Gacy, Fisher, & Gronowski, 2012; McDermott, Carolan, & Wickens, 2012; McDermott, Fisher, Carolan, Gronowski, Gacy, & Overstreet, 2012). Trainees learned to remotely drive a small unmanned ground vehicle along a predetermined route while looking for vehicles and identifying those vehicles as friendly, enemy, or neutral. Once a vehicle was spotted, the trainee hit a button to provide an alert of a potential threat. The trainee then classified the vehicle as friendly, enemy, or neutral once he or she was confident of the identification. This usually involved driving closer to the vehicle or manipulating the unmanned ground vehicle position to better situate the target vehicle in the camera's field of view. In sum, the task consisted of three component parts: mobility, target detection, and target identification. In the whole task, mobility was done simultaneously with detection (and was sometimes simultaneous with identification), but detection and identification were sequential.

Trainees were trained using a simulated vehicle and environment. They were later transitioned (far transfer) to the live unmanned ground vehicle in a live environment. Three different experimental training conditions varied whether training was part-task, whole-task, or both (i.e., part-task training of the three parts followed by whole-task practice scenarios) in the simulation environment. A fourth experimental condition involved part-task training in the simulation training environment, as well as part-task training in the live environment prior to starting transfer scenarios in the live environment. This fourth condition was the only condition to receive additional training in the live environment; the other conditions started directly on the whole task tests of transfer. In sum, the four experimental conditions were: (1) *part in simulation*, (2) *part and whole in simulation*, (3) *whole in simulation*, and (4) *part in simulation and live*. These conditions were chosen in order to make specific pairwise comparisons: the traditional comparison of part-task training to whole-task training, a comparison to examine the impact of part-task training prior to whole task training in simulation (as opposed to only whole task training in simulation), and a comparison to examine the potential advantage of additional part-task training in the live environment following part-task training in simulation (as opposed to no part-task training in the live environment, only in simulation).

In terms of the performance criteria, all conditions completed transfer scenarios in the live environment. Transfer to the live environment (and part-task training in the live environment for the condition that had additional training in the live environment) took place one week after initial training. Transfer was measured using two types of scenarios: (1) far transfer to a whole task in the live environment, which was repeated until the trainee achieved set performance criteria, and

(2) far transfer to a modified task in the live environment. In the modified task, there was no route and the trainee only had to detect enemy vehicles. Transfer performance criteria included: 100% detection and identification accuracy, with no collisions or wrong turns, and completing the transfer scenario within a specified time limit.

Findings. Two research questions addressed the relative advantages of part-task versus whole-task training in the simulation environment when both options are available and transfer is directly to the whole task in the live environment. The four experimental conditions were equally effective at getting 70-80% of the trainees to criterion-level transfer performance. There were no differences in transfer performance between the *part in simulation* group and the *whole in simulation* group. There was one marginally significant ($p = .065$) difference between the *whole in simulation* group and the *part and whole in simulation* group. On average, the *whole in simulation* group took fewer transfer scenarios to reach criteria than the *part and whole in simulation* group. Thus, there is some evidence that conducting only whole-task training in the training simulation was beneficial.

The most noteworthy findings concentrate on the usefulness of having additional part-task training in the live environment. Those with additional part-task training had significantly better mobility (i.e., fewer collisions) in the transfer scenarios. Trainees in this condition also reached transfer performance criteria in significantly fewer scenarios. Further, evidence suggests that this benefit was not due solely to the extra training time but also to the part-task method. Consistent with the benefits to training predicted by cognitive load theory, the additional part-task training *in the live environment* appears to have allowed trainees to focus on learning to timeshare the detection and mobility tasks during the transfer scenarios, while the other experimental conditions had to learn detection, mobility, *and* timesharing skills at the same time in the live environment. In addition, the additional live part-task training was more efficient than other conditions in terms of training time on the live robotic system. In the live part-task training, trainees used a part-task training application to practice detection and identification and then completed three mobility practice scenarios in which they teleoperated the unmanned vehicle. The three practice scenarios using the unmanned ground vehicle took an average of only 97 seconds but had a large payoff in terms of overall time saved. Those with additional part-task training spent an average of 46% less time on the live robotic system during the transfer scenarios than trainees in the other three conditions who started directly with a whole-task transfer scenarios. Thus, additional part-task training in the live environment rather than in the synthetic, training environment has the potential to save costs by reducing the amount of training time on the actual unmanned vehicle.

Worked Examples Experiment

In general, the research literature supports the effectiveness of the worked example training method for novice learning, as well as for structured problem-solving versus traditional problem-solving approaches (Sweller, van Merriënboer, & Paas, 1998; Van Gog & Rummel, 2010). However, the findings regarding the usefulness of worked examples are less understood for far transfer and for complex and unstructured problems such as those typically found in Army environments. Some research suggests that for novel problems requiring creative problem-solving, the reported benefit of the worked examples method tends to disappear (Sweller et al.,

1998). In contrast, other research found that worked examples can be effective for far transfer tasks (Atkinson, Derry, Renkl, & Wortham, 2000).

Our worked examples meta-analysis (Hutchins et al., 2013) investigated the benefit of worked examples for transfer. Overall, there was a significant transfer benefit for worked examples when compared to a control condition that did not use worked examples. As previously described, moderator analyses were conducted on a number of variables including the effect of task/skill type, transfer type, and trainee differences in ability and experience. Worked examples were effective for structured quantitative problems. Contrary to some previous studies (e.g., Sweller et al., 1998), the meta analysis also revealed a benefit of worked examples for non-quantitative and low structure problems. Worked examples were found to benefit identical and near transfer, but not far transfer. There were insufficient data to determine the moderating effects of trainee ability or experience on the usefulness of worked examples.

Accordingly, we designed an experiment to fill these gaps by investigating the usefulness of worked examples in a complex ill-structured task (for additional details, see McDermott, Carolan, & Gronowski, 2012). We extended the worked examples method to training complex decision making skills in real world problems in the form of unmanned vehicle route planning problems. The route planning task can vary in terms of how structured the problem is. Using such a planning problem task that can vary in task structure and in task difficulty was expected to provide more evidence for or against the effectiveness of worked examples in low structure task domains.

Experimental methodology. In the unmanned vehicle planning scenarios, trainees had to decide which unmanned vehicles to use in different geographic areas to satisfy multiple mission goals. This required an understanding of the tradeoffs of different assets in terms of mission suitability. The trainees did not have subject matter expertise on unmanned vehicles so we supplied a user-friendly manual that described the suitability of vehicles for different conditions. The scenarios did not have one right answer but required the trainee to consider multiple factors in making a decision. We manipulated whether or not a given trainee had access to a worked example. A second research question, not directly related to worked examples, addressed the use of media used in the training (i.e., training was conducted using a pencil–and-paper scenario or on a digital system).

In terms of performance criteria, performance was tested on a near transfer scenario immediately following training and far transfer scenarios a week later. No trainees had access to worked examples during the transfer scenarios. The near transfer scenario was more complex than the practice scenarios within the training; however, all trainees still performed on the media on which they were trained (the pencil–and-paper group used paper and the digital group performed on the digital system). The far transfer scenarios tested transfer to the digital system (for the group that learned on pencil-and-paper) and transfer to a more difficult problem that required adaptive problem solving (for both groups).

Findings. Contrary to the meta-analysis, worked examples did not provide a benefit over those without worked examples. Moreover, those *without* worked examples made significantly better planning choices both during the training and in the transfer scenarios. Those without

worked examples also considered a significantly higher number of factors (i.e., pieces of information) when choosing between unmanned vehicles in the transfer scenarios. The implication is that if consideration of a wide variety of factors is a critical part of the transfer task, then it may be detrimental to use worked examples during training; the rationale for this is that the worked examples may direct attention to certain factors and trainees may not learn to consider other factors not present in those examples. Thus, for complex problem solving and planning tasks, it may not be worth the development effort required to develop worked examples.

The transfer scenarios also tested how well the pencil-and-paper trained individuals transferred to the digital system. This group (vs. the digital group) took longer to complete the first transfer scenario on digital system, but their plan quality did not suffer. By the second transfer scenario, the pencil-and-paper trained group was just as quick as those who trained on the digital system, but their plan quality suffered. Therefore, trainees who learned on the digital system were successfully able to learn two tasks at once: the cognitive skills of making suitable plans and the procedural skill of implementing them in the digital system. Although it is a complex multi-faceted task with both cognitive and procedural aspects, trainees were able to master both in the whole task. One implication is that trainees do not need to learn general planning skills separate from learning to operate a digital system. They can successfully be learned in tandem.

Increasing Difficulty Experiment

This experiment sought to fill research gaps related to task difficulty sequencing identified in the increasing difficulty meta-analysis (see Wickens et al., 2013). In the meta-analysis, we examined constant difficulty (same level of difficulty across training), fixed/increasing difficulty (difficulty level increased across training on a set schedule) and adaptive difficulty (changes in difficulty level across training based on trainee performance). The meta-analysis found that adaptive increases in difficulty were more beneficial than fixed increases in difficulty. However, the pattern of results was less clear for constant difficulty versus adaptive difficulty. As such, this experiment was conducted to directly compare the usefulness of constant, fixed/increasing, and adaptive difficulty in the context of a complex cognitive task (for additional details, see McDermott, Gronowski, & Carolan, 2013).

Experimental methodology. The task used in this experiment was similar to the one used in the worked examples experiment described above. Trainees created plans on pencil-and-paper for unmanned vehicle allocation to multiple geographic areas, each with specific goals and anticipated enemy activity. Note that this task was more cognitively complex than previous published research studies investigating increasing difficulty.

There were four experimental training conditions that manipulated how trainees sequenced through training scenarios of different difficulty. *Constant* difficulty provided training scenarios at a consistently high difficulty level. *Fixed/increasing* difficulty started with a simple scenario and gradually increased difficulty up to a high level on a fixed schedule. Two *adaptive* difficulty conditions progressed through training scenarios according to an individual trainee's performance and decision process (i.e., information considered). Trainee performance determined the direction of the adaptation (poor performance decreased the difficulty, while proficient

performance increased difficulty) and the trainee's decision process determined the magnitude of the adaptation. In essence, by considering a high number of factors, the trainee moved up two difficulty levels instead of one. Likewise, if the trainee considered too few factors (in comparison to number of factors in the scenario), he/she moved down two levels of difficulty. *Adaptive Up* began with the lowest difficulty and *Adaptive Down* began with the highest difficulty. Trainees were matched to condition using a problem solving pretest. Training and transfer scenarios were examined to determine if training condition had an effect on performance.

Findings. We found that for a complex cognitive task, a constant (high) difficulty was superior to fixed/increasing difficulty on the far transfer tasks. The constant difficulty group had significantly higher plan quality in the second far transfer scenario than did the fixed/increasing group. The constant difficulty group also took the least amount of time to train; specifically, they took 14%, 57%, and 66% less time to train than the fixed/increasing group, adaptive up group, and adaptive down group, respectively. In comparing adaptive training to fixed/increasing difficulty, the fixed/increasing group considered significantly more factors in their decision making than the adaptive conditions for the first transfer scenario. The implication for military training is that complex cognitive tasks such as planning may not benefit from adaptive difficulty sequencing. Not having to create adaptive training modules also has the potential for cost savings in both training time and development time.

Adaptive Remediation Experiment

The increasing difficulty experiment had not shown a benefit of adapting difficulty when the task being trained was a complex planning task. However, we reflected that the increasing difficulty experiment did not specifically address the *type* of error made, which served as a trigger for adaptive changes. Therefore, we questioned whether adaptive remediation in response to the type of error made would aid transfer performance in a complex cognitive task. Accordingly, a follow-on experiment was conducted to study the impact of error prevention and adaptive remediation on transfer performance (for additional details, see McDermott, Gronowski, Carolan, & Fisher, 2013).

Experimental methodology. This experiment used the same general planning task used in the increasing difficulty experiment. The task involved creating plans for the allocation of unmanned vehicles to different geographic areas. However, while the prior experiment was conducted with pencil-and-paper, this experiment was conducted on a digital system. Note that all the trainees also previously completed the increasing difficulty experiment. This ensured that all trainees were experienced learners, allowing us to leverage the unmanned asset planning training they had already received.

Three experimental training conditions were developed to examine the impact of error prevention and adaptive remediation on performance. Error management training served as a basis of comparison to these two manipulations of error training and adaptive remediation. The three experimental conditions were as follows:
1. *Error Prevention* consisted of procedural training highlighting common errors associated with operating the digital system paired with procedural feedback during training. The

procedural feedback explained the particular steps (i.e., menu navigation, button clicks, factors to consider when choosing an asset) that needed to be taken to accomplish a task.

2. *Error Management* consisted of exploratory training in which the trainees were encouraged to explore the system and to use errors as an opportunity for learning. Trainees received feedback during the training which was conceptual; instead of telling trainees how to fix an error (as in the procedural feedback in the Error Prevention condition), trainees were given feedback on how the interfaces were organized to help the trainee find the solution (e.g., instead of telling them to click the RSTA button, conceptual feedback would state that viewing and classifying pictures was a different mode associated with Reconnaissance, Surveillance, and Target Acquisition).

3. *Adaptive Remediation* was identical to Error Management with the addition of remediation scenarios tailored to the particular errors made. The remediation scenarios offered additional practice opportunities on subtasks in which a trainee had made errors or could not complete. The remediation or practice opportunity was adaptive based on trainee performance.

In terms of performance criteria, transfer scenarios tested trainees' ability to adapt what was learned in the training to: (1) an error-prone transfer scenario that was designed with similar asset and location names and distractor assets, making it easy to make an error if the operator was not careful, (2) a transfer scenario that required the accomplishment of tasks that had not been specifically trained, and (3) a transfer scenario with a high tempo and conflicting goals.

Findings. There were no significant differences in transfer performance between the error prevention group with either the error management group or the adaptive remediation group. However, performance was better when trainees *did not* complete adaptive remediation. The error management group performed significantly more untrained tasks in the second transfer scenario than the adaptive remediation group, and also was more successful than the adaptive remediation group in rerouting unmanned assets in the third transfer scenario. Further, although the differences were not statistically significant, emerging trends suggested that the adaptive remediation group had more vehicle idle time, made worse asset choices, and were more likely to leave an unmanned asset in a vulnerable spot within line of sight of the enemy, as compared to the error management group. These differences are quite remarkable given that the only methodological difference between the error management and adaptive remediation conditions was that one group completed tailored remediation scenarios during training while the other did not. The poorer performance with adaptive remediation may be due to the fact that adaptive remediation encouraged trainees to focus on single tasks and therefore they spent less time exploring and learning the entire system. Also, the fact that the trainees were experienced learners may explain why the cognitive load reducing adaptive remediation training method was not effective. In sum, the use of error management (non-remediation) training has the potential to increase transfer performance, as well as cost savings due to reduced training time and less development time for training developers in comparison to adaptive remediation training.

Interpersonal Skills Literature Review

Although the present research effort focused on complex cognitive skills, we also conducted a focused literature review and synthesis on interpersonal skills to examine its viability as a future research direction (for the full review, see Hutchins, McDermott, Carolan, Gronowski, Fisher, & DeMay, 2013). One of the particular goals of this literature review was to identify topics that may warrant a closer examination via meta-analysis. In other words, how large is the existing body of research evidence and have any prior meta-analyses and/or literature reviews been conducted on the topic? Two literature searches were conducted: (1) a systematic search of "interpersonal skills" training research literature and (2) a survey of the literature on six particular interpersonal skills. These particular interpersonal skills were identified as being potentially relevant to current and future Army operations and included: relationship building skill, nonverbal communication skill, negotiation skill, assertive communication skill, active listening skill, and conflict resolution skill. Note that a subset of the papers identified in these literature reviews were summarized and included within the TARGET database.

During our review of the broad interpersonal skills literature, we sought to organize the types of interpersonal skills being researched and the key training interventions used. We identified and defined 28 individual interpersonal skills and numerous taxonomies of interpersonal skills (for examples, see Carpenter & Wisecarver, 2004; Doo, 2006). Note, that Carpenter and Wisecarver (2004) was the only validated taxonomy we identified. In terms of key training interventions used, results from two meta-analyses suggested that overall, interpersonal skills training is relatively effective at improving interpersonal skills (Arthur et al., 2003; Klein, 2009). Further, a literature review by Klein, DeRouin, and Salas (2006) found evidence that it is more beneficial for interpersonal skills training to focus on specific, optimal social skills as opposed to focusing on general skills such as sensitivity or insight. Moreover, our review of training, assessment, and measurement methods suggested that while most training programs targeting the development of interpersonal skills are multi-method programs, the core training technique used is behavioral modeling training (e.g., instruction, demonstration, role-play/practice, and feedback). Indeed, behavior modeling has been shown to be beneficial in two meta-analyses (Klein, 2009; Taylor et al., 2005), across a variety of task/skill types and for several training outcomes including cognitive outcomes, skill-based outcomes, and job behavior.

Surprisingly, our broad literature review revealed that little is known about the impact of virtual (vs. in-person) practice as a training technique, and more generally, how well interpersonal skill training (regardless of training method) *transfers* to real life situations. More research is needed to determine when and where other training strategies, besides behavioral modeling, will be effective for improving specific interpersonal skills. Additional research is also needed to examine the transfer effectiveness beyond the training environment and into workplace and real life situations.

With respect to the six skill-specific surveys of experimental literature, our review provided insight into the current state of interpersonal skills research, identified research needs, and identified several topic areas that may warrant a meta-analysis. The review looked for differences in training method effectiveness and the outcome of far transfer in particular. Note that we found an increase in experimental publications over the past decade for all six interpersonal skills examined, yet a general lack of research examining interpersonal skill transfer. The current state of the literature and a sample of key findings are described below:

- For the skill of 'relationship building,' no prior literature reviews or meta-analyses were found; however, we concluded that this skill category is potentially too broad of a construct. The two reports that focused on relationship building used behavioral modeling as the training method and showed improvement in relationship building skills at the end of the training (Durlach, Wansbury & Wilkinson, 2008; Schlundt, Quesenberry, Pichert & Lorenz, 1994).

- The skill of 'nonverbal communication' generated a handful of literature reviews (see Gladstein, 1974; Klinzing & Gerada-Aloisio, 2004; Rosenthal, Wadsworth, Russell, Mathew, Elfenbein, Sanchez-Burks & Ruark, 2009) and meta-analyses (see Klinzing & Tisher, 1986). There has been a surge of both encoding and decoding experimental training research in nonverbal communication in the past two decades that could be synthesized. The most successful form of nonverbal communication skill training utilized skill practice.

- There were two more recent reviews of 'negotiation skills' (see Logan, 2001; Tsay, & Bazerman, 2009) and two meta-analyses (see Stuhlmacher & Citera, 2005; Zetik & Stuhlmacher, 2002), but very little synthesis has been performed with a focus on training effectiveness. There has been a lack of variety in the training methods used to train negotiation skills. Studies found that performance was moderated by individual differences of personality and cultural background (see Elfenbein, Curhan, Eisenkraft, Shirako, & Brown, 2009)

- There is a fairly large body of experimental research on training 'assertive communication,' but the one existing meta-analysis (Shatz, 1984) is dated and at least 40 studies have been published since 1984. Recent research suggests that training programs for assertive communication that included role-play practice and feedback were superior to training with (1) lecture and demonstration and (2) training with lecture only.

- The experimental research on 'active listening' has steadily increased over the last five decades; however, no literature reviews or meta-analyses were found, making it a prime candidate for qualitative and quantitative synthesis of the literature. The training effectiveness of active listening programs with role-play was superior to those without role-play.

- There has been a surge in experimental attention on 'conflict resolution,' but no meta-analyses and only one review has been conducted (see Boulter, Von Bergan, Miller, & Wells, 1995). This makes conflict resolution another prime candidate for meta-analyses in the future. The most effective training programs for conflict resolution skill involved modeling, practice and feedback, and the research suggests that training effectiveness is moderated by individual differences and prior experiences.

Development of Algorithms

Algorithms were developed to quantify the relationships between the training methods, performance, and the various moderating factors. These algorithms can be used to perform tradeoff analyses for different combinations of training methods. The algorithms make the research findings from this project available to the Army training, development, and research communities. Specifically, the algorithms allows trainers and researchers to: (1) systematically explore the meta-analytic evidence base to identify training methods that would be effective for a particular set of circumstances for acquiring cognitive skills and (2) add new research studies to this evidence base along with providing real time computational updates (discussed further in the next section). Steps involved in the algorithm development are briefly summarized below (for full details, see Hutchins, Carolan, Plott, McDermott, & Orvis, 2014).

The methods of Borenstein et al. (2009) were used within a single study to calculate effect sizes from raw study-level information (e.g., means, correlations). The Bornstein et al. methods were also used to compile effect sizes *across* multiple research studies. As aforementioned, the effect size is a statistical concept that measures the strength of the relationship between two variables (Preacher & Kelly, 2012). We focused on Hedges' g (Hedges & Olkin, 1985) as a standardized measure of effect size between the treatment group (i.e., experimental group receiving the training method) and the control group (i.e., experimental group receiving no training or a lesser degree of the given training method). Implementation of the computational effect size algorithms in TARGET included procedural steps for transforming a number of different types of raw study-level data (e.g., descriptive statistics, t-test statistics, and F-test statistics) to standardized individual effect size statistics, which were then used to summarize the overall effect size across a set of research studies.

The algorithms allow for each research study to contribute multiple comparisons (or effects) between treatment and control groups, each requiring computation of an effect size estimate for the magnitude of the difference in performance between the two groups. As an example, a study could contribute a comparison between the treatment and control for near transfer performance and for far transfer performance. An attribute coding scheme, based on the meta-analyses, was implemented to support the process of examining key moderators of training effectiveness for each training method. This included, for example, the trainee characteristics, task/skill types, outcome measures (e.g., knowledge acquisition, near transfer), and the training method-specific moderators (e.g., concurrent vs. sequential tasks in part-task training). This coding scheme also supports the capability for a user to filter, analyze, and display subsets of effect size data. For example, a user could choose to only view data related to trainee experience. In this example, the data would be displayed for two categories: low experience and high experience trainees.

Note that there were several combinations of moderating variables for which there is no extant research available (e.g., the impact of trainee experience on the effectiveness of part-task training of perceptual skills) and thus the aforementioned algorithms could not be directly employed to generate effect sizes. Accordingly, we also defined and implemented an innovative extrapolation process for estimating such effect size statistics. Such effect size estimates are based on extrapolating from a subset of moderator variables that <u>does</u> include the target variables – that

is, from the nearest "parent" (e.g., the impact of trainee experience on the effectiveness of part-task training *over all task/skill types,* with 'all task/skill types' being the parent of 'perceptual skills'). The extrapolation strategy was designed to maintain effect size relationships between moderator variables while minimizing algorithm complexity.

The algorithms offer several potential benefits to the Army. As implemented in TARGET (described below), they can be used to estimate the expected costs or benefits of the six training methods on performance, for various combinations of task/skill type, trainee characteristics, and performance outcomes. For example, the algorithms help answer questions such as: How do worked examples impact performance on a psychomotor task? Does this effect depend on the experience or skill of the trainee?

The algorithms could also be adapted for use in other human performance models, such as Improved Performance Research Integration Tool (IMPRINT). IMPRINT was developed by the U.S. Army Research Laboratory, Human Research and Engineering Directorate (ARL-HRED) to support Manpower and Personnel Integration (MANPRINT) and HSI analyses. IMPRINT is a modeling tool designed to help assess the interaction of Soldier and system performance. With IMPRINT, users can gain useful information about processes that might be too expensive or time-consuming to test in the real world. Adding the training effects algorithms from this research into IMPRINT would allow users to determine how different training methods might impact performance and predict which training methods (or combination of methods) will result in the most effective performance. Either of these uses can assist program managers with training design. The algorithms can provide the basis for cost benefit analyses of different training methods and may enable program managers to make decisions concerning the amount of training that system operators and maintainers should receive, as well as what basic types of training methods should be developed to support this training.

TARGET Tool

A key goal of the current research effort was to assist training developers and researchers in better understanding the relative effectiveness of different training methods for acquiring cognitive skills. As such, a comprehensive research database was generated that included our meta-analytic findings of the six training methods, the qualitative summaries of these six methods, as well as the interpersonal skills literature reviewed. To ensure the valuable findings from this research database would be easily consumable by training developers and researchers, a training effectiveness tool was developed, called TARGET (which stands for Training Aide: Research and Guidance for Effective Training).

TARGET was designed to assist users in making evidence-based decisions concerning the most suitable training method(s) to use depending various moderators of interest: in particular, depending on the task/skill type(s) being trained, types of trainees that would participate in the intended training, and/or the performance outcomes sought. TARGET provides query-based searches of the database, as well as outputs textual summaries, numerical summaries, and graphical representations of the relationships between different training methods and performance outcomes; this is completed using a user-friendly graphical user interface (GUI) and the underlying algorithms described in the prior section. TARGET is also updateable as additional

training research is generated so that the database/tool stays current with state-of-the-art research developments.

TARGET is a web-based tool; the website is publicly accessible at **http://bldr-webtest.alionscience.com/Target/**. Anyone can access TARGET, after completing a brief, free registration. Below, we briefly describe the four main components of TARGET: Explore Tasks, Explore Methods, Explore Documents, and Add a New Document. The tool is designed to allow easy navigation among the four components. For additional details on TARGET's capabilities, please visit the TARGET website and/or review the TARGET User Guide (Plott & Hutchins, 2013).

Explore Tasks

Explore Tasks is a visualization that allows users to quickly explore accumulated research evidence relating to different task/skill types. The assumption is that a training developer or researcher will have information about the type of task/skill that needs to be trained. This component serves as a starting point for those users to visualize which training method(s) have been successfully used to train the task/skill type of interest, and which ones should be avoided as they represent a 'cost' to training performance outcomes. The visualization displays the task at the center of the screen surrounded by the six training methods. The closer a given training method is to the task at the center of the screen, the more *evidence* there is that this training method benefits performance for the task of interest (see Appendix as an example). The Explore Tasks component of TARGET provides users with a range of interactive features for viewing accumulated evidence by task/skill type in order to identify which training method(s) are likely to be useful in training a particular task/skill type(s).

Explore Methods

The Explore Methods component graphically displays results of the meta-analyses for a single training method at a time. The visualization allows users to explore various effect size information on the effectiveness of a given training method (e.g., Learner Control) and allows the user to drill down to examine the impact of different moderators on the relationship between this method and training performance. These moderators include task/skill type, task difficulty, trainee characteristics, outcome criteria, and training method-specific moderators, among others. As aforementioned in the meta-analysis section of this report, the training method-specific moderators are unique to the implementation of each training method. For example, one part-task training specific moderator is the degree of task concurrence; that is, the degree to which the subtasks are completed concurrently and require timesharing.

Figure 2 shows an example of the Explore Methods visualization using data from the Learner Control training method. The visualization conveys the findings graphically so that in-depth statistical knowledge is not required. For example, the vertical dotted line down the middle of the table indicates neutral or no effect and either side is labeled as "cost" and "benefit" allowing the user to interpret if there was a cost or benefit to training performance. The skinny diamond in the top row represents the *overall* effect size information for Learner Control (compiled across all research studies examining this training method). In this example, there is
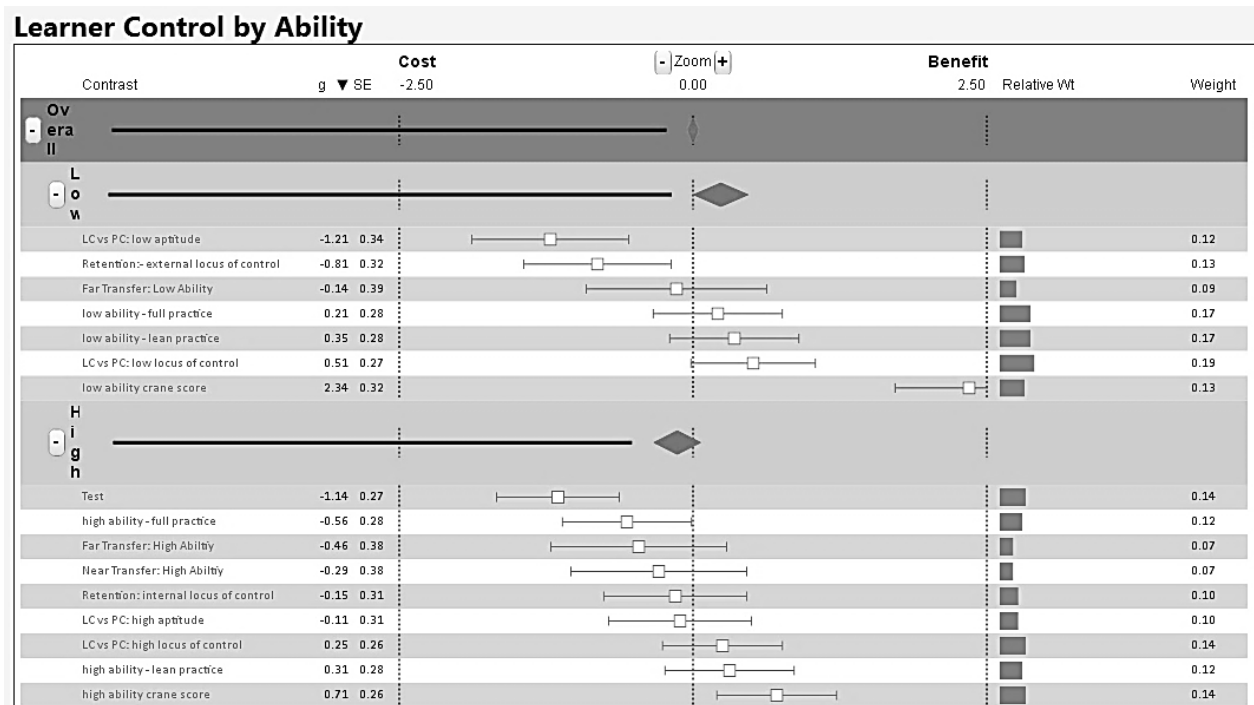
neither a benefit nor cost to training performance when using the Learner Control training method. In this screen shot, the user has also chosen to examine how the trainee characteristic of 'ability' moderates the effects of this training method on performance. The effect size data for low and high ability are summarized in the diamonds next to the respective category of low or high. The visualization suggests there is a small overall benefit in using this training method for lower ability trainees, while there is a small overall cost for trainees higher in ability. Finally, under each ability category is a list of the individual effects (i.e., from the individual research studies) that contribute to the category. The relative weight bars on the right of the display show how much each research study influences the diamond for that category. Users interested in further details about the evidence can follow the hyperlinked title of the study effect (e.g., LC vs PC: low aptitude) to access more specific information about the particular research study.

Explore Documents

The Explore Documents component links users to the searchable TARGET database of over 500 research studies housed in the tool. Using a custom-built advanced search capability, users are able to search the training literature in the tool by traditional paper features (author, publication source, publication year), as well as training study attributes (e.g., training method, task/skill type, outcome criterion). For each research study, users can view reference information, a qualitative summary of the study's methods and findings, the relevant attributes (i.e., training method used, training task information, trainee characteristics examined, performance outcomes examined), and effect size statistics if applicable.

Add New Document

The Add New Document component links users to a wizard that provides step-by-step instructions on how to enter information/statistics from a new research study into the TARGET database. This functionality was developed so that the TARGET research database can be kept up-to-date as new research evidence becomes available, as well as so the user can compare the findings of this new study with findings derived from the research literature already housed in the tool. Users can enter reference information, qualitatively summarize methods and findings, choose study attributes, and enter statistical data. The new study information is visible to other users.

Learner Control by Ability

Contrast	g	SE	Weight
Overall			
Low			
LC vs PC: low aptitude	-1.21	0.34	0.12
Retention:- external locus of control	-0.81	0.32	0.13
Far Transfer: Low Ability	-0.14	0.39	0.09
low ability - full practice	0.21	0.28	0.17
low ability - lean practice	0.35	0.28	0.17
LC vs PC: low locus of control	0.51	0.27	0.19
low ability crane score	2.34	0.32	0.13
High			
Test	-1.14	0.27	0.14
high ability - full practice	-0.56	0.28	0.12
Far Transfer: High Ability	-0.46	0.38	0.07
Near Transfer: High Ability	-0.29	0.38	0.07
Retention: internal locus of control	-0.15	0.31	0.10
LC vs PC: high aptitude	-0.11	0.31	0.10
LC vs PC: high locus of control	0.25	0.26	0.14
high ability - lean practice	0.31	0.28	0.12
high ability crane score	0.71	0.26	0.14

Figure 2. Example of Explore Methods forest plot

Summary

In summary, the overarching goal of this 4-year research effort was to develop evidence-based guidelines for the relative effectiveness of six different training methods for acquiring and transferring cognitive skills in complex task domains. To accomplish this overall goal, we focused on four main research objectives.

First, we sought to summarize the current state of the training effectiveness literature and identify research gaps. Accordingly, we conducted a broad literature search to gather evidence on the effectiveness of various training methods. From the literature review, we narrowed the possible training method options to those identified as most suitable for cognitive skills. Based on this literature review, we conducted six comprehensive meta-analyses in order to generate estimates of the effectiveness of the following training methods: training wheels, scaffolding, part-task training, increasing difficulty, exploratory learning, and learner control. We also focused heavily on identifying the impact of moderators on a given training method's effectiveness. That is, we were not only interested in identifying whether a method had an overall benefit (or lack thereof), but in identifying for whom specifically that benefit existed and if this benefit varied according to the training performance outcome(s) of interest and type of cognitive task/skill trained.

Our second research objective was to conduct a series of research experiments to help fill several identified research gaps from the meta-analyses. A total of five experiments were conducted that included several common design elements. For example, the experiments involved training complex Army-relevant tasks, examined the impact of trainee characteristics, and assessed various types of transfer performance. The results contribute to a more complete body of

knowledge concerning training complex cognitive tasks. These experimental results were used to update the meta-analytic results, which were then used to implement the final two research objectives.

Our third research objective was to develop algorithms to quantify the relationships between the six training methods, performance, and various moderating factors. These algorithms can be used to perform tradeoff analyses for different combinations of training methods. The algorithms make the research findings from this project available to the Army training, development, and research communities, allowing users to systematically explore training methods and design components that would be effective for a particular set of circumstances for acquiring cognitive skills. These algorithms can be applied in a variety of applications from decision support tools for training developers to input for human performance models to analyze the impact of different training or technological approaches. For example, the algorithms could also be adapted for use in IMPRINT to determine how different training methods might impact performance.

Finally, to ensure these research findings and algorithms would be easily consumable by various users, our fourth research objective was to develop a user-friendly graphical user interface tool, called TARGET. This tool contains key elements from the literature review, meta-analyses, and experimentation. TARGET summarizes the cognitive skill training research and identifies the conditions under which a particular method is more or less effective. TARGET contains several visualization tools, such that in-depth statistical knowledge is not required to benefit from this tool. Accordingly, training developers (with varying levels of expertise) can easily use TARGET's evidence-based recommendations to identify the most effective training method given a set of desired factors/conditions. These capabilities within TARGET are facilitated by the underlying algorithms.

In addition to training developers, researchers may benefit from TARGET by better understanding the state of the training literature, including any possible gaps in the field's understanding of effective methods. Military service program managers can also use this information to direct future research to fill identified gaps or investigate currently inconclusive findings. The capabilities represented in the training tool can serve a number of potential future applications as well, such as expanding the database to include new training methods (e.g., behavior modeling) or task/skill types (e.g., interpersonal skills) and/or adapting the tool's architecture to a different literature domain beyond training.

References

Alliger, G., Tannenbaum, S., Bennett, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology, 50*, 341-358.

Alvarez, K., Salas, E., & Garofano, C. M. (2004). An integrated model of training evaluation and effectiveness. *Human Resource Development Review, 3*, 385-416.

Arthur, W., Jr., Bennett, W., Jr., Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology, 88*, 234-245.

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. W. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70,* 181–214.

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology, 41*, 63-105.

Barnett, S. M. & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*, 612-637.

Bayraktar, S. (2002). A meta-analysis of the effectiveness of computer assisted instruction in science education. *Journal of Research on Technology in Education, 34*, 173-188.

Bell, B. S., & Kozlowski, S. W. (2008). Active learning: Effects of core training design elements on self-regulatory processes, learning, and adaptability. *Journal of Applied Psychology, 93*, 296-316.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* West Sussex, UK: John Wiley & Sons, Ltd.

Boulter, A., Von Bergan, C. W., Miller, M. J., & Wells, D. (1995). Conflict resolution: An abbreviated review of current literature with suggestions for counselors. *Education, 116*, 93-97.

Blume, B., Ford, J., Baldwin, T., & Huang, J. (2009). Transfer of training: A meta analytic review. *Journal of Management, 36*, 1065-1105.

Carolan, T., McDermott, P. L., Hutchins, S., Wickens, C. D., & Belanich, J. (2011). Investigating the impact of training on performance. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference.* Arlington, VA: National Defense Industrial Association.

Carolan, T. F., McDermott, P. L., Wickens, C. D., Fisher, A., & Gronowski, M. (under review). Guidance and exploratory learning: Application to transfer and retention in a procedural learning task. *Cognitive Engineering and Decision Making Journal*.

Carolan, T. F., Wickens, C. D., Hutchins, S. & Cumming, J. (2014). Costs and benefits of more learner freedom: Meta-analyses of exploratory and learner control training methods. *Human Factors*.

Carpenter, T. D., & Wisecarver, M. M. (2004). *Identifying and validating a model of interpersonal performance dimensions* (Technical Report 1144). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Carroll, J. M. (1990). *The Nurnberg funnel: Designing minimalist instruction for practical computer skills*. Cambridge, MA: MIT Press.

Clark, R. E. (2009). How much and what type of guidance is optimal for learning from instruction? In S. Tobias & T. Duffy (Eds.), *Constructivist theory applied to instruction: Success or failure?* (pp. 158-183). New York: Taylor and Francis.

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development, 42*, 21-29.

Clark, R. E. (1984). Reconsidering research on learning from media. *Review of Educational Research, 53*, 445-459.

Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology, 85*, 678-707.

Corbalan, G., Kester, L., & van Merriënboer, J. J. G. (2011). Learner-controlled selection of tasks with different surface and structural features: Effects on transfer and efficiency. *Computers in Human Behavior, 27*, 76-81.

Damos, D., & Wickens, C. D. (1980). The identification and transfer of time-sharing skills. *Acta Psychologica, 46*, 15–39.

Doo, M. Y. (2006). A problem in online interpersonal skills training: Do learners practice skills? *Open Learning*, *21*, 263-272.

Doolittle, P. (2010). The effects of segmentation and personalization on superficial and comprehensive strategy instruction in multimedia learning environments. *Journal of Educational Multimedia and Hypermedia, 19*, 159-175.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and Educational Psychology. *Psychological Science in the Public Interest, 14*, 4-58.

Durlach, P. J., Wansbury, T. G., & Wilkinson, J. G. (2008). *Cultural awareness and negotiation skills training: Evaluation of a prototype semi-immersive system.* Paper presented at the Army Science Conference. Orlando, FL.

Elfenbein, H., Curhan, J., Eisenkraft, N., Shirako, A., & Brown, A. (2009). *Why are some negotiators better than others? Opening the black box of bargaining behaviors.* Social Science Research Network Working Paper. Available at SSRN: http://ssrn.com/abstract=1336257.

Fontana, F. E., Mazzardo, O., Furtado, O., Jr, & Gallagher, J. D. (2009). Whole and part practice: A meta-analysis. *Perceptual & Motor Skills, 109*, 517-530.

Ford, J. K., & Weissbein, D. A. (1997). Transfer of training: An updated review and analysis. *Performance Improvement Quarterly, 10*, 22-41.

Frederiksen, J. & White, B. (1989). An approach to training based upon principled task decomposition. *Acta Psychologica, 71*, 89-146.

Frese, M., Bodbeck, F., Heinbokel, T., Mooser, C., Schleiffenbaum, E., & Thiemann, P. (1991). Errors in training computer skills: On the positive function of errors. *Human-Computer Interaction*, *6*, 77-93.

Gladstein, G. A. (1974). Nonverbal communication and counseling/psychotherapy: A review. *The Counseling Psychologist*, *4*, 34-52.

Goettl, B. P., & Shute, V. J. (1996). Analysis of part-task training using the backward-transfer technique. *Journal of Experimental Psychology: Applied, 2*, 227-249.

Gully, S., & Chen G. (2010). Individual differences, attribute-treatment interactions, and training outcomes. In S. W. J. Kozlowski & E. Salas (Eds.), *Learning training, and development in organizations* (pp. 3 – 64). New York: Routledge.

Healy, A. F., & Bourne, L. E. (2011). Applied cognitive psychology. In P. R. Martin, F. M. Cheung, M. C. Knowles, M. Kyrios, J. B. Overmier, & J. M. Prieto (Eds.), *IAAP handbook of applied psychology*. Oxford, UK: Wiley-Blackwell.

Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* London: Academic Press.

Heimbeck, D., Frese, M., Sonnentag, S., & Keith, N. (2003). Integrating errors into the training process: The function of error management instructions and the role of goal orientation. *Personnel Psychology, 56*, 333-361.

Hutchins, S. D., Carolan, T. F., Plott, B. M., McDermott, P. L., & Orvis, K. A. (2014). *Development of the 'TARGET' training effectiveness tool and underlying algorithms specifying training method – performance outcome relationships.* (Research Note 2014-

02). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Hutchins, S., McDermott, P. L., Carolan, T. F., Gronowski, M. R., Fisher, A., & DeMay, M. (2013). *Interpersonal skills summary report*. (Research Note 2013-03). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Hutchins, S., Wickens, C. D., Carolan, T., & Cumming, J. (2013). The influence of cognitive load on transfer with error prevention training methods: A meta-analysis. *Human Factors, 55*, 854-874.

Ivancic, B., & Hesketh, K. (2000). Learning from error in a driving simulation: Effects on driving skill and self-confidence. *Ergonomics, 43*, 1966–1984.

Kalyuga, S. (2009). Instructional designs for the development of transferable knowledge and skills: A cognitive load perspective. *Computers in Human Behavior, 25*, 332-338.

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review, 23*, 1-19.

Kamouri, A. L., Kamouri, J., & Smith, K. H. (1986). Training by exploration: Facilitating the transfer of procedural knowledge through analogical reasoning. *International Journal of Man-Machine Studies, 24*, 171-192.

Keith, N., & Frese, M. (2005). Self-regulation in error management training: Emotion control and metacognition as mediators of performance effects. *Journal of Applied Psychology, 90,* 677-691.

Keith, N. & Frese, M. (2008). Effectiveness of error management training: A meta-analysis. *Journal of Applied Psychology, 93*, 59-69.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75-86.

Klein, C.R. (2009). *What do we know about interpersonal skills? A meta-analytic examination of antecedents, outcomes, and the efficacy of training*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3357877)

Klein, C., DeRouin, R. E., & Salas, E. (2006). Uncovering workplace interpersonal skills: A review, framework, and research agenda. In G. P. Hodgkinson & J. K. Ford (Eds.), *International review of industrial and organizational psychology,* (Vol. 21, pp. 79–126). New York: John Wiley & Sons.

Klinzing, H. G. & Gerada-Aloisio, B. (2004). *Intensity, variety, and accuracy in nonverbal cues and de-/encoding: Two experimental investigations.* Paper presented at the Annual Meeting of the American Educational Research Association. San Diego, CA.

Kozlowski, S. W .J., Toney, R. J., Mullins, M. E., Weissbein, D. A., Brown, K. G., & Bell, B. S. (2001). Developing adaptability: A theory for the design of integrated embedded training systems. In E. Salas (Ed.), *Advances in human performance and cognitive engineering research* (Vol. 1, pp. 59-123). Amsterdam: JAI/Elsevier Science.

Kraiger, K., Ford, J., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*, 311-328.

Kraiger, K. & Jerden, E. (2007). A meta-analytic investigation of learner control: Old findings and new directions. In S. M. Fiore & E. Salas E. (Eds.), *Towards a science of distributed learning* (pp. 65-90). Washington, DC: American Psychological Association.

Landers, R. N. (2009). *Traditional, web-based, and hybrid instruction: A comparison of training methods.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3358633).

Lintern, G., & Wickens, C. D. (1991). Issues for acquisition in transfer of timesharing and dual-task skills. In D. Damos (Ed.), *Multiple-task performance* (pp. 123–138). London: Taylor & Francis.

Lu, S., Wickens, C., Hutchins, S., Sarter, N. & Sebok, A. (2013) Supporting interruption management and multimodal interface design: Three meta-analyses of task performance as a function of interrupting task modality. *Human Factors, 55*, 697 – 724.

Logan, M. H. (2001). *What hinders or facilitates successful crisis negotiation* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. NQ71498).

Mane, A. M., Adams, J. A., & Donchin, E. (1989). Adaptive and part-whole training in the acquisition of a complex perceptual-motor skill. *Acta Psychologica, 71*, 179–196.

Mayer, R. E. (2004). Should there be a three-strike rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist, 59*, 14-19.

Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93*, 390-397.

McDaniel, M. A., & Schlager, M. S. (1990). Discovery learning and transfer of problem-solving skills. *Cognition and Instruction, 7*, 129-159.

McDermott, P. L., Carolan, T. F., Fisher, A., Gronowski, M. R., & Gacy, M. (2013). *Transferring from the simulator to a live robotic environment: The effectiveness of part-task and whole-task training*. (Technical Report 1331). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

McDermott, P. L., Carolan, T., Gacy, M., Fisher, A., & Gronowski, M. R. (2012). Cost effective approaches to training reconnaissance tasks. *Proceedings of the Infotech@Aerospace Conference.* Reston, VA: American Institute of Aeronautics and Astronautics.

McDermott, P. L., Carolan, T., & Gronowski, M. R. (2012). Application of worked examples to unmanned vehicle route planning. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference.* Arlington, VA: National Defense Industrial Association.

McDermott, P. L., Carolan, T., & Wickens, C. D. (2012). Part-task training methods in simulated and realistic tasks. *Proceedings of the 56th Annual Meeting of the Human Factors and Ergonomics Society*. Boston, MA: Human Factors and Ergonomics Society.

McDermott, P. L., Fisher, A., Carolan, T., Gronowski, M. R., Gacy, M., & Overstreet, M. (2012). Transfer from a simulation environment to a live robotic environment: Are certain demographics better? *Proceedings of the 7th ACM/IEEE International Conference on Human-Robot Interaction.* Washington, DC: Institute of Electrical and Electronics Engineers.

McDermott, P. L., Gronowski, M. R., & Carolan, T. (2013). Adaptive training strategies: Considering the decision process as well as the outcome. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference.* Arlington, VA: National Defense Industrial Association.

McDermott, P. L., Gronowski, M. ., Carolan, T. F., & Fisher, A. (2013). Error training and adaptive remediation: The impact on transfer performance in a complex planning task. *Proceedings of the 57th Annual Meeting of the Human Factors and Ergonomics Society.* San Diego, CA: Human Factors and Ergonomics Society.

Means, B., Toyama, Y., Murphy. R., Bakia, M., & Jones, K. (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online-learning studies*. Washington, DC: U.S. Department of Education.

Metzler-Baddeley, C., & Baddeley, R. J. (2009). Does adaptive training work? *Applied Cognitive Psychology, 23*, 254-266.

Orvis, K.A., Fisher, S. L., & Wasserman, M. E. (2009). Power to the people: Using learner control to improve trainee reactions and learning in web-based instructional environments. *Journal of Applied Psychology, 94*, 960-971.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*, 1–4.

Paas, F., & van Gog, T. (2009). Principles for designing effective and efficient training of complex cognitive skills. *Reviews of Human Factors and Ergonomics, 5*, 166-194.

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences, 13,* 423–451.

Plott, B., & Hutchins, S. (2014). *Training aide: Research and guidance for effective training - User guide.* (Research Product 2014-02). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Preacher, K. J., & Kelly, K. (2012). On effect size. *Psychological Methods, 17*, 137-152.

Pridemore, D. R., & Klein, J. D. (1991). Control of feedback in computer-assisted instruction. *Educational Technology Research and Development, 39*, 27-32.

Quiñones, M. A. (1997). Contextual influences on training effectiveness. In M. A. Quiñones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Applications of psychological research* (pp. 177-199). Washington, DC: American Psychological Association

Rey, G., & Buchwald, F. (2011). The expertise reversal effect: Cognitive load and motivational explanations. *Journal of Experimental Psychology: Applied, 17*, 33–48.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181–210.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Revised Edition). Beverly Hills, CA: Sage.

Rosenthal, D. B., Wadsworth, L. A., Russell, T. L., Mathew, J., Elfenbein, H. A., Sanchez-Burks, J., & Ruark, G. A. (2009). *Training soldiers to decode nonverbal cues in cross-cultural interactions* (Research Note 2009-12). Arlington, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Sanders, W. R. (1999). *Digital procedural skill retention for selected M1A2 tank intervehicular information systems* (Research Report 1432). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Schlundt, D. G., Quesenberry, L., Pichert, J. W., & Lorenz, R. A. (1994). Evaluation of a training program for improving adherence promotion skills. *Patient Education and Counseling*, *24*, 165-173.

Schmidt, R. A. & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.

Schnackenberg, H. L., & Sullivan, H. J. (2000). Learner control over full and lean computer-based instruction under differing ability levels. *Educational Technology Research and Development, 48*, 19-35.

Seamster, T., Redding, R., & Kaempf, G. (1997). *Applied cognitive task analysis in aviation*. Aldershot, UK: Avebury Aviation.

Shatz, M. A. (1984). *Assertiveness training: A meta-analysis of the research findings* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 8325006).

Shen, C.-Y., & Tsai, H.-C. (2009). Design principles of worked examples: A review of the empirical studies. *Journal of Instructional Psychology, 36*, 238–244.

Shyu, H. Y. (1993). Who should control the interactivity of instruction? Issues in learner control. *Journal of Educational Media & Library Sciences, 31*, 1-16.

Sitzmann, T., Kraiger, K., Stewart, D., & Wisher, R. (2006). The comparative effectiveness of web-based and classroom instruction: A meta-analysis. *Personnel Psychology, 59*, 623-664.

Smith, E. M., Ford, J. K., & Kozlowski, S. W. J. (1997). Building adaptive expertise: Implications for training design strategies. In M. A. Quinones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace* (pp. 89–118). Washington, DC: American Psychological Association.

Snow, R. (1989). Aptitude-treatment interaction as a framework for research on individual differences in learning. In P. Ackerman, R.J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences*. New York: W.H. Freeman.

Stuhlmacher, A. F., & Citera, M. (2005). Hostile behavior and profit in virtual negotiation: A meta-analysis. *Journal of Business and Psychology, 20*, 69-93.

Sweller, J. (1988). Cognitive load during problem-solving: Effects on learning. *Cognitive Science, 12*, 257–285.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous and germane cognitive load. *Educational Psychology Review, 22,* 123–138.

Sweller, J., van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296.

Tang, Z. (2004). *Learner control in an interactive learning environment.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database.

Tannenbaum, S. F., Mathieu, J. E., Salas, E. & Cannon-Bowers, J. A. (1991). Meeting trainees' expectations: The influence of training fulfillment on the development of commitment, self-efficacy, and motivation. *Journal of Applied Psychology, 76*, 759-769.

Taylor, P. J., Russ-Eft, D. F., & Chan, D. W. L. (2005). A meta-analytic review of behavior modeling training. *Journal of Applied Psychology, 90*, 692–709.

Taylor, R. S. (2005). *Informal science learning: Influences of explanatory elaboration and learner control on knowledge acquisition.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database.

Tsay, C-J., & Bazerman, M. H. (2009). *A decision-making perspective to negotiation: A review of the past* (Working paper 10-002*). Cambridge, MA:* Harvard Business School.

U.S. Department of the Army (2010). *The operations process* (Field Manual 5-0). Retrieved from https://armypubs.us.army.mil/doctrine/index.html.

U.S. Department of the Army, U.S. Army Training and Doctrine Command (2011). *The U. S. Army learning concept for 2015* (TRADOC Pamphlet 525-8-2). Retrieved from www-tradoc.army.mil/tpubs/pams/tp525-8-2.pdf

van Gog, T., Paas, F., & Sweller, J. (2010). Cognitive load theory: Advances in research on worked examples, animations, and cognitive load measurement. *Educational Psychology Review, 22*, 375–378.

van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review, 22*, 155–174.

van Merriënboer, J. J. G., Kester, L., & Paas, F. (2006). Teaching complex rather than simple tasks: balancing intrinsic and germane load to enhance transfer of learning. *Applied Cognitive Psychology, 20,* 343–352.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering psychology and human performance* (4th Ed). Boston, MA: Pearson.

Wickens, C. D., Hutchins, S., Carolan, T., & Cumming, J. (2013). The effectiveness of part-task training and increasing-difficulty training strategies: A meta-analysis approach. *Human Factors, 55*, 461-470.

Wightman, D., & Lintern, G. (1985). Part-task training for tracking and manual control. *Human Factors, 27*, 267–284.

Wood, R. E. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes, 37*, 60-82.

Zetik, D. C., & Stuhlmacher, A.F. (2002). Goal setting and negotiation performance: A meta-analysis. *Group Processes & Intergroup Relations, 5*, 35-52.

Appendix. Example of an *Explore Tasks* graphical visualization in TARGET